PCT/CA 00/00966
29 SEPTEMBER 2000 (29.09.00)

REC'D 11 OCT 2000
WIPO PCT

PA 305969

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

4

September 25, 2000

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

CA 00/00966

APPLICATION NUMBER: *60/159,122*
FILING DATE: *October 13, 1999*

10/069304

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

W. Montgomery

W. MONTGOMERY
Certifying Officer

PTO/SB/16(6-95)
Approved for use th[redacted] 04/11/98. OMB 0651-0037
Patent and Trademark Office, U.S. D[redacted]TMENT OF COMMERCE

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2).

Docket Number 2846/0002		Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
SILVA	Nancy	F.	66 Woodhaven Crescent Richmond Hill, Ontario Canada L4E 3T2
GORING	Daphne	R.	2240 Taylors Orchard Mississauga, Ontario Canada L5B 2T3

TITLE OF THE INVENTION (280 characters max)

PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES

CORRESPONDENCE ADDRESS

Gervas W. Wall
Deeth Williams Wall
150 York Street, Suite 400
Toronto

STATE	Ontario	ZIP CODE	M5H 3S5	COUNTRY	Canada

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification	*Number of Pages*	59	☒ Small Entity Statement
☒ Drawings	*Number of Sheets*	22	☐ Other (specify) Application cover sheet

METHOD OF PAYMENT *(check one)*

☒ A check or money order is enclosed to cover the Provisional filing fees ☐ The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number	PROVISIONAL FILING FEE AMOUNT ($)	$75.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No

☐ Yes. The name of the U.S Government agency and the Government Contract number are:

Respectfully Submitted.

SIGNATURE _____ Date 12/10/99

TYPED or PRINTED NAME GERVAS W. WALL REGISTRATION NO. 35,766 (if appropriate)

☐ Additional inventors are being named on separately numbered sheets attached hereto

DEETH WILLIAMS WALL
Barristers & Solicitors Patent & Trade Mark Agents

National Bank Building
150 York Street, Suite 400
Toronto, Ontario
Canada M5H 3S5

Douglas N. Deeth
Gervas W. Wall
Victor V. Butsky
Michael Erdle
Anne Thomas
Adrian M. Kaplan
Victor Krichker
Shereen Hamdy

Amy-Lynne P. Williams
A. Marijo Coates
Eileen M. McMahon

Diane M. LaCalamita
Gordon S. Jepson
Noel Courage
Ailbe Flynn

Telephone: (416) 941-9440
Facsimile: (416) 941-9443

Gervas W. Wall
Direct Dial: (416) 941-9001
email: gwall@dww.com

File No.: 2864/0002

October 12, 1999

DELIVERED BY HAND

Assistant Commissioner for Patents
Patent and Trademark Office
Correspondence and Mail Division
Crystal Plaza Building 2, Room 1A03
2011 South Clark Place
Arlington, VA 22202

Dear Sirs:

Re: New U.S. Provisional Application
Title: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
Applicants & Inventors: SILVA, Nancy F. & GORING, Daphne R.

We apply in the of names Nancy F. Silva and Daphne R. Goring for a provisional patent application entitled Proline-rich Extensin-like Receptor Kinases.

We enclose the following documents:

1. provisional application cover page;
2. statement claiming small entity status; and
3. patent application.

The (U.S.) $75.00 filing fee is included in our firm cheque.

If you have any questions, please call Gervas W. Wall at (416) 941-9001 or Noel Courage at (416) 941-0954.

Yours truly,

Gervas W. Wall
Registration No. 35,766

GWW/NAC/rc
Encl.

Proline-rich Extensin-like Receptor Kinases

FIELD OF THE INVENTION

The invention relates to nucleic acid molecules and polypeptides involved in plant defense, and more particularly increasing plant resistance to wounding and pathogens.

BACKGROUND OF THE INVENTION

Receptor mediated signal perception and transduction in response to external stimuli are essential for growth and developmental processes of multicellular organisms (Mu et al., 1994). These extensively well characterized processes in animal systems involve receptor protein kinase molecules comprised of an extracellular signal perception domain, a hydrophobic transmembrane domain attached to an intracellular domain that possesses kinase activity (Horn et al., 1994). In general, transmembrane signaling by receptor protein kinases requires binding of an appropriate ligand to the extracellular domain which induces receptor dimerization and alters the activity of the intracellular catalytic domain. This promotes phosphorylation of specific substrates thereby initiating a protein kinase signaling cascade (Ullrich and Schlessinger, 1990). The majority of animal receptor protein kinases isolated to date contain tyrosine-specific kinase domains (Ullrich and Schlessinger, 1990), however, the transforming growth factor beta (TGF-beta) receptor (Lin et al., 1992) and the activin receptor (Dijke et al., 1993) possess kinase domains with serine/threonine phosphorylation activity.

Intracellular communication is also essential for the growth and development of higher plants. The extensive knowledge of cell surface receptor signaling in animal systems has resulted in the isolation of several genes predicted to encode receptor-like protein kinases (RLKs). The characterized members of the RLK family share highly homologous catalytic domains with consensus sequences indicative of serine/threonine autophosphorylation activity, yet the extracellular domains of these receptors are very divergent (Braun and Walker, 1996). Five different classes of plant receptor-like protein kinases have therefore been identified according to amino acid sequence similarity in the extracellular domains of these genes. The first class of receptor kinases, designated the S-domain class, have distinct extracellular domains homologous to the S-locus glycoprotein (SLG) (Nasrallah and Nasrallah, 1993). S-domain

receptor kinases have several distinguishing features such as ten conserved cysteine residues located proximal to the transmembrane domain in addition to other conserved residues implicated in the proper folding of the extracellular domain (Walker, 1994). Among this class of receptor kinases are the S-locus receptor kinases (SRKs) of Brassica expressed exclusively in reproductive tissues and implicated along with SLGs in controlling the sporophytic self-incompatibility response which normally inhibits self-pollination (Stein et al., 1991; Goring and Rothstein, 1992). Other receptor-like kinases of this type are represented in Arabidopsis by ARK1, ARK2, ARK3 (Tobias et al., 1992; Dwyer et al., 1994), in maize by ZmPK1 (Walker and Zhang, 1990) and by OsPK10 in rice (Zhao et al., 1994). The diversity in patterns of expression among members of the S-domain class shows that these plant receptor kinases are involved in mediating a variety of cellular signaling processes (Walker, 1994).

Another class of plant receptor kinases is the leucine-rich repeat (LRR) group which encodes proteins with extracellular domains containing 20-25 imperfect repeats of a 24 amino acid leucine-rich motif involved in peptide ligand recognition, cell adhesion and implicated in mediating protein-protein interactions (Braun and Walker, 1996; Wang et al., 1998). This class of plant receptor kinases include proteins such as CLAVATA1 (Clark et al., 1997) which is involved in regulating meristem and flower development in Arabidopsis, as well as proteins functioning in gamete development such as PRK1 of Petunia (Mu et al., 1994). The LRR class is represented in Arabidopsis by other receptor kinases such as ERECTA (Torii et al., 1996) which has been shown to be essential for proper plant and organ elongation, BRI1, a receptor involved in brassinosteroid signal transduction (Li and Chory, 1997), as well as TMK1 (Chang et al., 1992) and RLK5 (Walker, 1993) which may have more general roles in cellular signaling as suggested by their ubiquitous expression patterns in a variety of vegetative and reproductive tissues. Xa21, another member of the LRR class, has been implicated in pathogen recognition by providing resistance in rice to *Xanthomonus oryzae* pv. *oryzae* (Song et al., 1995).

The lectin-like class of plant receptor kinases is represented only in *Arabidopsis thaliana* by Ath.lecRK1 (Hervé et al., 1996) and LRK1 (Swarup et al., 1996). The extracellular domain of these receptor kinases share sequence similarity with lectins which are known carbohydrate

binding proteins and implicated in the transduction of oligosaccharide signals in plant cellular communication processes (Hervé et al., 1996).

The two remaining classes of plant receptor kinases isolated in *Arabidopsis thaliana* include proteins with extracellular domains containing epidermal growth factor (EGF)-like motifs found in many proteins involved in extracellular interactions (WAK1; Kohorn et al., 1992) as well as thaumatin-like domains homologous to PR5 proteins involved in pathogenesis (PR5K; Wang et al., 1996).

Plants remain very vulnerable to wounding and pathogens despite these advances. There is a need to identify other polypeptides, that help to protect plants. There is also a need for transgenic plants which overexpress these polypeptides and which have increased resistance to wounding and pathogens.

SUMMARY OF THE INVENTION

The invention relates to nucleic acid molecules and polypeptides involved in plant defense, and more particularly increasing plant resistance to wounding and pathogens.

We isolated a cDNA clone designated PERK1 (Proline-rich Extensin-like Receptor Kinase 1) which encodes a receptor kinase in *Brassica napus*. We define a new class of PERK plant receptor kinases characterized by an extracellular domain rich in proline sharing sequence similarity to the extensin family of cell wall proteins. PERK1 is induced by both wounding and chemical elicitors which mimick a pathogen attack, showing a role for PERK1 in mediating a plant's defense response to mechanical and biological attack. Similar PERK nucleic acid molecules and polypeptides are found in other plants and cells.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments will be described in relation to the drawings in which:

Figure 1. Nucleotide and Deduced Amino Sequence. [SEQ ID NOS.: 1 & 2]

(A) Figure 1 shows the nucleic acid molecule of [SEQ ID NO.: 1] and the amino acid sequence of [SEQ ID NO.: 2].

In a preferred embodiment, the figure shows the DNA sequence corresponding to the PERK1 transcript, with the predicted amino acid sequence presented as a single-letter code below the nucleotide sequence. Numbers to the left refer to nucleotide sequence and the 5' and 3' untranslated regions are presented in lower case letters. Potential N-glycosylation site (Asn-x-Ser/Thre) are indicated by dots above the Asn residues, and the extensin signature $(Ser\text{-}Pro)_4$ pentapeptide motif present in the extracellular domain is denoted in boldface type. The predicted membrane spanning region is marked by a solid underline. As defined by Hanks and Quinn (1991), the catalytic domain has been subdivided into 11 subdomains marked by dashed underlines and superscript roman numerals. The amino acids in redface type represent residues that are absolutely conserved, whereas greenface type represent groups of conserved amino acids. The two regions marked by double underlines represent consensus sequences common amongst serine/threonine kinases.

(B) Structural features of the PERK1 polypeptide. A Kyte hydropathy plot (Kyte and Doolittle, 1982) of the predicted amino acid sequence generated by DNAsis® software (Hitachi Software, San Bruno, CA) is shown, where increased hydrophobicity is denoted by positive values. The domains of the PERK1 protein are illustrated below. ECD, extracellular domain; TM, transmembrane domain.

(C) Shows the nucleic acid molecule of [SEQ ID NO.: 3] and the amino acid sequence of [SEQ ID NO.: 4]. These sequences are identical to nucleotides 1 to 1944 of [SEQ ID NO:1] and the corresponding amino acid sequence but they also include 5' and 3' untranslated nucleotide regions. **(D)** Shows the nucleic acid molecule of [SEQ ID NO.: 3] including the 5' and 3' untranslated regions. The start and stop codons are underlined and in red print. The nucleotide numbers correspond to the entire sequence and not only to the coding region. In a preferred embodiment, this is the sequence of PERK1

(E) Figure 1 shows the amino acid sequence of [SEQ ID NO.: 5] including amino acids corresponding to the 5' and 3' untranslated regions. The starting methionine is in red print and the stop codon is indicated by a red print asterisk. It will be clear to those skilled in the art that one may only use the portion of the amino acid sequence between the start and stop codons.

Figure 2. Genomic DNA Southern Blot Analysis of PERK1.

Genomic DNA (5 micrograms) isolated from *Brassica napus* leaf tissue was digested with the indicated restriction enzymes, blotted and hybridized with a partial 1.5 kb PERK1 cDNA probe under varying conditions of stringency. DNA markers are indicated in kilobases.

(A) Genomic DNA gel blot analysis under low stringency conditions.

(B) Genomic DNA gel blot analysis under high stringency conditions.

Figure 3. Expression of PERK1 cDNA.

(A) RNA gel blot analysis of PERK1 transcripts in poly(A)$^+$ mRNA extracted from various *Brassica napus* tissues. The blot hybridized with a partial 1.5 kb PERK1 cDNA probe detected a full length transcript of ~2.2kb (open-face triangle).

(B) The blot was subsequently probed with cyclophilin as an internal control for even loading (bold-face triangle).

(C) Ethidium bromide stain of gel indicates relatively equal amounts of poly(A)$^+$ mRNA were loaded.

Figure 4. Wound-Inducible Accumulation of PERK1 mRNA in *Brassica napus* Leaf and Stem Tissue.

(A) Fully expanded leaves were wounded by punching out discs around the perimeter of the leaf blade. Wounds mimic injury inflicted on plants in the field as a result of insect attack or other mechanical damage. Total RNA was extracted at various time intervals after treatment, subjected to Northern blot analysis and probed with full length PERK1 cDNA (open-face triangle).. The blot was reprobed with cyclophilin used as an internal control for even loading (open-face triangle). The graph represents the expression profile of PERK1 in response to wounding corrected against levels of cyclophilin expression. Control unwounded leaf tissue represented by 0 hr time point.

(B) Northern blot showing a time-course induction of PERK1 mRNA accumulation in wounded stem tissue. Total RNA harvested at the indicated time points was blotted and hybridized against the full length PERK1 coding sequence (open-face triangle). The cyclophilin loading control

(bold-face triangle) was used to normalize levels of PERK1 mRNA accumulation represented graphically. Control unwounded stem tissue represented by 0 hr time point.

Figure 5. Effects of 50 micromolar Methyl Jasmonate (MeJA) on PERK1 mRNA Accumulation in Treated *Brassica napus* Leaf and Stem Tissue.

(A) *Brassica napus* plants were thoroughly sprayed with a 50 micromolar MeJA solution, and leaf tissue subsequently harvested at different time intervals after treatment. Total RNA prepared from treated leaf tissue was subjected to Northern blot analysis and probed with full length PERK1 cDNA (open-face triangle). Control plant (0 hr) was treated with the carrying solution minus the chemical inducer (0.1% [$^{v}/_{v}$] ethanol for MeJA). The blot was reprobed with cyclophilin used as an internal control for even loading (bold-face triangle). The graph represents a corrected profile for the levels of PERK1 mRNA accumulation in response to treatment with MeJA normalized against levels of cyclophilin expression.

(B) Northern blot showing a time-course induction of PERK1 mRNA accumulation in MeJA treated stem tissue. Total RNA harvested at the indicated time points was blotted and hybridized against the full length PERK1 coding sequence (open-face triangle). The cyclophilin loading control (bold-face triangle) was used to normalize levels of PERK1 mRNA accumulation represented graphically.

Figure 6. Effects of 4mM Salicylic Acid (SA) on PERK1 mRNA Accumulation in Treated *Brassica napus* Leaf and Stem Tissue.

(A) *Brassica napus* plants were thoroughly sprayed with a 4mM SA solution, and leaf tissue subsequently harvested at different time intervals after treatment. Total RNA prepared from treated leaf tissue was subjected to Northern blot analysis and probed with full length PERK1 cDNA (open-face triangle). Control plant (0 hr) was treated with the carrying solution minus the chemical inducer (5mM phosphate buffer, pH7). The blot was reprobed with cyclophilin used as an internal control for even loading (open-face triangle). The graph represents a corrected profile for the levels of PERK1 mRNA accumulation in response to treatment with SA normalized against levels of cyclophilin expression.

(B) Northern blot showing a time-course induction of PERK1 mRNA accumulation in SA treated stem tissue. Total RNA harvested at the indicated time points was blotted and hybridized against the full length PERK1 coding sequence (open-face triangle). The cyclophilin loading control (bold-face triangle) was used to normalize levels of PERK1 mRNA accumulation represented graphically.

Figure 7. Proposed pathway mediating PERK1 expression in response to wounding, MeJA and SA treatments.

Figure 8. Western blots.

(A) Represents a Western blot performed on the bacterially expressed extracellular domain fusion protein using a T7 monoclonal antibody which recognizes the T7 epitope on the bacterially expressed fusion protein .

(B) Western blot to confirm that the extracellular domain fusion protein is targeted to inclusion bodies.

Figure 9. Western blot conducted on the bacterially expressed PERK1 protein to determine whether the full length protein is insoluble and forms inclusion bodies.

Figure 10.

(A) Western blot to confirm the induction and purification of both the wild-type (bold-face star) and mutated (bold-face arrow) catalytic domain fusion proteins using an anti-MBP antibody.

(B) represents a kinase assay performed on affinity purified wild-type and mutated fusion proteins incubated in the presence of γ-^{32}PdATP.

Figure 11. Shows sequence identity of PERK1 to polypeptides from the Arabidopsis genome sequencing project.

Figure 12. Shows the nucleic acid molecule of [SEQ ID NO.: 6] and the amino acid sequence of [SEQ ID NO.: 7].

In a preferred embodiment, the figure shows the figure shows the sequence of the predicted Arabidopsis gene - Accession number AAC98010

A) Genomic Sequence. The predicted open reading frame is underlined. The start codon (ATG)

and stop codon (TGA) are double underlined.

B) Translation of the predicted open reading frame. The transmembrane domain is underlined.

Figure 13. Shows the nucleic acid molecule of [SEQ ID NO.: 8] and the amino acid sequence of [SEQ ID NO.: 9].

In a preferred embodiment, the figure shows the sequence of the predicted Arabidopsis gene - Accession number AAD15491

A) Genomic Sequence. The predicted open reading frame is underlined. The start codon (ATG) and stop codon (TGA) are double underlined.

B) Translation of the predicted open reading frame. The transmembrane domain is underlined.

Figure 14. Shows the nucleic acid molecule of [SEQ ID NO.: 10] and the amino acid sequence of [SEQ ID NO.: 11].

In a preferred embodiment, the figure shows the sequence of the predicted Arabidopsis gene - Accession number CAA18823

A) Genomic Sequence. The predicted open reading frame is underlined. The start codon (ATG) and stop codon (TGA) are double underlined.

B) Translation of the predicted open reading frame. The transmembrane domain is underlined.

Figure 15. Shows the nucleic acid molecule of [SEQ ID NO.: 12] and the amino acid sequence of [SEQ ID NO.: 13].

In a preferred embodiment, the figure shows the sequence of the predicted Arabidopsis gene - Accession number CAA18590

A) Genomic Sequence. The predicted open reading frame is underlined. The start codon (ATG) and stop codon (TAG) are double underlined.

B) Translation of the predicted open reading frame. The transmembrane domain is underlined.

DETAILED DESCRIPTION OF THE INVENTION

In this study, we report the isolation and preliminary characterization of PERK nucleic acid molecules and polypeptides, and in particular PERK1 cDNA which encodes a novel

receptor-like protein kinase in *Brassica napus*. PERK polypeptides represent a novel class of receptor kinases in higher plants.

Protein kinases play important roles in plant defense (Zhou et al., 1995; Usami et al., 1995; Suzuki and Shinhi, 1995). Significant homology is shared between the extracellular domain of PERK1 and both extensin and proline rich proteins. PERK1 mediates plant responses to mechanical wounding (ie. insect attack) and pathogen attack. PERK polypeptides preferably include a a catalytic domain. PERK polypeptides are also preferably signaling molecules associated with the cell wall via their extensin-like extracellular domain and involved in the transduction of extracellular stimuli (eg. wounding, pathogen attack) into an intracellular response through a cytoplasmic kinase domain, thereby bridging the cell wall - plasma membrane continuum.

In general, plants challenged by mechanical wounding or pathogen attack induce rapid expression of genes (ie. proteinase inhibitor *(pin)* and pathogenesis related *(PR)* genes respectively) that are expressed locally as well as systemically in unaffected parts of the plant (Yang et al., 1997). Increased levels of extensin transcripts as a result of mechanical wounding have been well established in many other systems (Sauer et al., 1990; Shirsat et al., 1996). For example, in *Brassica napus* leaf and stem tissue, wound induction of PERK1 mRNA accumulation is a very rapid response detectable within 15 min following injury (Figure 4). Despite the well established involvement of MeJA (the methyl ester of the plant growth regulator jasmonic acid (JA)) in the signal transduction pathway regulating gene activation upon wounding, processes occurring immediately after wounding remain poorly characterized in terms of additional components that may also participate in wound signaling (Titarenko et al., 1997). Steady state levels of PERK1 mRNA remain unaffected to exogenously applied MeJA (Figure 5) which shows that the inducibility of PERK1 by wounding occurs via a MeJA-independent pathway (Figue 7). Studies conducted by Titarenko et al. (1997) addressing the role of JA in mediating wound responses support the existence of multiple distinct wound signal transduction pathways. Exogenously applied JA was able to induce only a subset of wound responsive genes in Arabidopsis which ultimately resulted in a stronger systemic accumulation in wounded plants. Conversely, a second set of wound responsive genes showing a stronger

induction locally in wounded tissue showed no substantial accumulation upon JA treatment. In conjunction with the pattern of PERK1 mRNA accumulation in response to wounding and MeJA, it appears that plants respond to wounding by two distinct wound signal transduction pathways: one which does not require JA and is primarily responsible for gene activation in the vicinity of the wound site and the other which involves JA perception and activates gene expression both locally and systemically to the wound site (Titarenko et al., 1997).

Many of the inducible defense responses are not exclusive to mechanical wounding but are also initiated by pathogen attack. The similarity between responses to wounding and pathogen attack are not surprising since mechanical damage often precedes pathogen infection and conversely, mechanical damage may often result from a pathogen or insect attack (Truernit et al., 1996). Salicylic acid has been implicated in having an important role in the signal transduction pathway leading to systemic acquired resistance (SAR) (Penninckx et al., 1996). Steady state levels of PERK1 mRNA also accumulated in *B. napus* leaf and stem tissue upon exogenous application of 4mM SA (Figure 6). Collectively, the profiles of PERK1 mRNA accumulation in response to wounding, MeJA and SA are not entirely surprising. PERK1 induction is rapid in response to wounding (Figure 4) and the lack of PERK1 transcript accumulation in response to MeJA (Figure 5) shows a pathway for wound mediated induction of PERK1 that is independent of MeJA (Figure 7). The pronounced and rapid induction of PERK1 in response to exogenous SA (Figure 6) supports other studies showing that SA is known to inhibit wound responsive genes that are regulated by a MeJA-dependent pathway (Peña-Cortés et al., 1993; Doares et al., 1995). Therefore, it is unlikely that both MeJA and SA would induce PERK1 mRNA accumulation given that these pathways are known to be antagonistic (Peña-Cortés et al., 1993). Nevertheless, the rapid induction of PERK1 during these treatments shows a role early on in a plant's defense signaling pathway.

Characterization of PERK1

Genomic Southern blot analysis under low and high stringency conditions revealed that PERK1 is a single copy gene in the *Brassica* genome (Figure 2). PERK1 is ubiquitously expressed at high levels in root, stem, petal and pistil and is less abundant in leaf tissue (Figure 3A).

The deduced amino acid sequence of PERK1 shows that it is a transmembrane receptor kinase with a distinct extracellular, transmembrane and cytoplasmic domain (Figure 1). The extracellular domain of PERK1 shows sequence similarity to plant cell wall proline-rich proteins and extensins which comprise a family of hydroxyproline-rich glycoproteins (HRGPs). Extensins are particularly abundant proteins in plant cell walls and are very rich in proline and serine as well as in combinations of valine, tyrosine, lysine and/or histidine residues. The distinctive characteristic of dicot extensins is their repetitive $(Ser\text{-}Pro)_4$ pentapeptide blocks. Although extensins are synthesized as soluble precursors, the majority of proline residues are hydroxylated and both the hydroxylated proline as well as the serine residues of these proteins are glycosylated by post-translational modifications (Cassab, 1998). When secreted to the plant cell wall, extensins become rapidly insoluble, presumably due to the formation of covalent isodityrosine bridges (Cassab, 1998). Although extensins have been proposed to be structural cell wall proteins and important in development, they have also been directly implicated in plant defense against mechanical wounding (Shirsat et al., 1996) and pathogen attack (Corbin et al., 1987; Showalter, 1993). The catalytic domain of PERK1 possesses all of the invariant residues necessary for kinase activity and sequence similarity in subdomains VI and VIII to amino acid consensus sequences characteristic of serine/threonine kinases shows a role for PERK1 in plant signal transduction (Hanks and Quinn, 1991). Given the similarity of PERK1 in the extracellular domain to the extensin family of cell wall proteins, PERK1 can detect changes to the cell wall through mechanical damage or pathogen attack and then pass the signal onto the cell. The cell can then respond to the attack with its defence mechanisms.

PERK nucleic acid molecules and proteins also have sequence identity and similarity to proline rich proteins as well. About 40% of PERK1's extracellular domain is comprised of proline. Below is a list of the Arabidopsis clones and their respective proline composition (Table below).

Table 1

% Prolines in Extracellular Domains

Gene	% Proline in ECD
PERK1	56/137 = 41%
CA18590	105/279 = 38%
AAC98010	85/246 = 34%
CAA18823	51/179 = 28%
AAD15491	36/149 = 24%

It is possible to use all or part of a proline rich domain from an extensin or a proline rich protein (or similar regions) to replace all or part of PERK1's extracellular domain. In summary, PERK1 is a unique plant protein in *Brassica napus* involved in wound and pathogen response which physically links the cell wall and plasma membrane. PERK1 is involved in the general perception and subsequent transduction of a wound and/or pathogen stimulus, ultimately triggering a plant's defense mechanisms and conferring broad protection against such stimuli. Preliminary characterization showed that levels of PERK1 mRNA accumulate rapidly in response to wounding and SA. Further characterization of PERK1 induction with respect to changes in levels of phosphorylation provides additional evidence for the unequivocal role of PERK1 in plant defense. Furthermore, transgenic analysis of plants expressing altered levels of PERK1 confirms the involvement of PERK1 in wound and pathogen signaling. PERK1 polypeptides and nucleic acid molecules may be isolated from the Brassicaceae including Arabidopsis, *Brassica napus*, *Brassica rapa, Brassica juncea, Brassica oleracea*, and other plants such as potato, tomato, tobacco, cotton, carrot, petunia, sunflower, strawberries, spinach, lettuce, rice, soybean, corn, wheat, rye, barley, sorgum and alfalfa.

PERK Nucleic Acid Molecules and Polypeptides

The invention relates to PERK nucleic acid molecules and polypeptides which increase wounding resistance and pathogen resistance in cells and plants. These polypeptides preferably include an extracellular domain, a transmembrane domain and a cytoplasmic domain. The cytoplasmic domain preferably includes a region with kinase activity. The kinase activity is involved in cellularsignaling. The PERK nucleic acid molecules which encode PERK polypeptides are particularly useful in producing transgenic plants which have increased wounding and pathogen resistance compared to a wild type plant.

It will also be apparent that there are polypeptide and nucleic acid molecules from other organisms, such as those listed previously that are similar to PERK polypeptides and nucleic acid molecules. PERK polypeptides are useful in increasing wounding and pathogen resistance in a cell, preferably a plant cell, because they include extensin-like and proline-rich domains (this refers to a plurality of domains including multiple proline residues, which are preferably similar to those found in extensins), such as SPPPP, SPP and PPP which are capable of being hydroxylated in response to wounding or pathogen attack. Once hydroxylated, extensins become rapidly insoluble which strengthens the cell wall and in response to pathogen attack helps agglutinate or prevent spread of the pathogen to neighbouring plant cells.

The PERK nucleic acid molecules and polypeptides, as well as their role in plants were not known before this invention. The ability of these compounds to increase wounding and pathogen resistance of transgenic host cells (particularly plant cells) and transgenic plants compared to wild type cells and plants was unknown.

All nucleotides and polypeptides which are suitable for use in the methods of the invention, such as the preparation of transgenic host cells or transgenic plants, are included within the scope of the invention. Genomic clones or cDNA clones are preferred for preparation of transgenic cells and plants.

In a preferred embodiment, the invention relates to cDNAs encoding PERK polypeptides from *Brassica napus*. The cDNA sequence and the corresponding amino acid sequence for

PERK1 is presented in Figure 1. The invention also includes splice variants of the nucleic acid molecules as well as polypeptides produced from the molecules.

Characterization of Nucleic Acid Molecules and Polypeptides

In one variation, the invention includes DNA sequences (and the corresponding polypeptide) including at least one of the sequences shown in figure 1 in a nucleic acid molecule of preferably about: less than 1000 base pairs, less than 1250 base pairs, less than 1500 base pairs, less than 1750 base pairs, less than 2000 base pairs, less than 2250 base pairs, less than 2500 base pairs, less than 2750 base pairs or less than 3000 base pairs.

The coding region of the PERK1 nucleic acid molecule is as follows:

Table 2

Nucleic Acid Molecule	Start Nucleotide [brackets show corresponding amino acid nos.]	End Nucleotide [brackets show corresponding amino acid nos.]
PERK1 (coding region only)	**1 (1)**	**1944 (648)**
PERK1 Extracellular Domain	**1 (1)**	**411 (137)**
PERK1 Transmembrane Domain	**412 (138)**	**480 (160)**
PERK1 Cytoplasmic Domain	**481 (161)**	**1944 (648)**
PERK1 Kinase region	**Same as cytoplasmic domain**	**Same as cytoplasmic domain**

It will be apparent that these may be varied, for example, by shortening the 5' untranslated region or shortening the nucleic acid molecule so that the 3' end nucleotide is in a different position.

The discussion of the nucleic acid molecules, sequence identity, hybridization and other aspects of nucleic acid molecules included within the scope of the invention is intended to be applicable to either the entire nucleic acid molecule in figure 1 or the coding region of this molecules, shown in Table 2. One may use the entire molecule in figure 1 or only the coding region. Other possible modifications to the sequence are apparent.

Southern Blot Analysis shows that PERK1 is present as a single copy gene in *Brassica*. A Northern blot showed that PERK1 polypeptide was expressed in all tissues examined (root, shoot (shoot includes leaves and stems), flower and petal). It is highly expressed in the stem, petal and root and to a lower extent in the leaf and pistil.

The PERK1 Nucleic Acid Molecule and Polypeptide are Conserved in Plants

Sequence Identity

This is the first isolation of a nucleic acid molecule encoding a PERK polypeptide from plant species. Nucleic acid sequences having sequence identity to the PERK1 sequence are found in other species of Brassica such as *Brassica rapa, Brassica juncea, and Brassica oleracea* as well as other plants such as Arabidopsis, potato, tomato, tobacco, cotton, carrot, petunia, sunflower, strawberries, spinach, lettuce, rice, soybean, corn, wheat, rye, barley, sorgum and alfalfa. Sequences from *Brassica napus* and other plants are collectively referred to as "PERK" nucleic acid sequences and polypeptides. We isolate PERK nucleic acid molecules from plants having nucleic acid molecules that are similar to those in *Brassica napus,* such as beet, tomato, rice, cucumber, radish and other plants including Arabidopsis, potato, tobacco, cotton, carrot, petunia, sunflower, strawberries, spinach, lettuce, soybean, corn, wheat, rye, barley, sorgum and alfalfa. and using techniques described in this application. The invention includes methods of isolating these nucleic acid molecules and polypeptides as well as methods of using these nucleic acid molecules and polypeptides according to the methods described in this application, for example those used with respect to PERK1.

Table 3 below shows several sequences with sequence identity and sequence similarity to the PERK1 polypeptide. Where polypeptides are shown, a suitable corresponding DNA encoding the polypeptide will be apparent. These sequences code for polypeptides similar to

portions of PERK1 polypeptide. The sequences in Table 3 are useful to make probes to identify full length sequences or fragments (from the listed species or other species). One skilled in the art would be able to design a probe based on a polypeptide or peptide fragment. The invention includes nucleic acid molecules of about: 10 to 50 nucleotides, 50 to 200 nucleotides, 200 to 500 nucleotides, 500 to 1000 nucleotides, 1000 to 1500 nucleotides, 1500 to 1700 nucleotides, 1700 to 2000 nucleotides, 2000 to 2500 nucleotides or at least 2500 nucleotides and which include all or part of the sequences (or corresponding nucleic acid molecule) in Table 3. The invention also includes a nucleic acid molecule including the sequences in Table 3 which encodes peptides and polypeptides of about: 10 to 50 amino acids, 50 to 200 amino acids, 200 to 500 amino acids, 500 to 750 amino acids or at least 750 amino acids. Possible modifications to these sequences will also be apparent. The polypeptide and nucleic acid molecules are also useful in research experiments or in bioinformatics to locate other sequences. The nucleic acid molecules and polypeptides preferably provide PERK activity.

Table 3

Organism	Accession No.
Arabidopsis thaliana	AAC 98010 (Figure 12)
Arabidopsis thaliana	AAD 15491 (Figure 13)
Arabidopsis thaliana	CAA 18823 (Figure 14)
Arabidopsis thaliana	CAA 18590 (Figure 15)

The invention includes the nucleic acid molecules from other plants as well as methods of obtaining the nucleic acid molecules by, for example, screening a cDNA library or other DNA collection with a probe of the invention (such as a probe comprising at least about: 10 or preferably at least 15 or 30 nucleotides of PERK1 and detecting the presence of a PERK nucleic acid molecule. Another method involves comparing the PERK1 sequences to other sequences, for example using bioinformatics techniques such as database searches or alignment strategies, and detecting the presence of a PERK nucleic acid molecule or polypeptide. The invention includes the nucleic acid molecule and/or polypeptide obtained according to the methods of the

invention. The invention also includes methods of using the nucleic acid molecules, for example to make probes, in research experiments or to transform host cells or make transgenic plants. These methods are as described below.

The polypeptides encoded by the homologous PERK nucleic acid molecules in other species will have amino acid sequence identity to the PERK1 sequence. Sequence identity may be at least about: >20%, >25%, >28%, >30%, >35%, >40%, >50% to an amino acid sequence shown in figure 1 (or a partial sequence thereof). Some polypeptides may have a sequence identity of at least about: >60%, >70%, >80% or >90%, more preferably at least about: >95%, >99% or >99.5% to an amino acid sequence in figure 1 (or a partial sequence thereof). Identity is calculated according to methods known in the art. Sequence identity (nucleic acid and protein) is most preferably assessed by theFasta 3 program , using the following default parameter settings: gap penalty (open) = -12 (protein) –16 (DNA), gap penalty (extension) = -2 (protein) –4 (DNA) , protein weight matrix = BLOSUM 62. (The reference for FASTA 3 is W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448,and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA" Methods in Enzymology 183:63- 98). . The invention also includes modified polypeptides from plants which have sequence identity at least about: >20%, >25%, >28%, >30%, >35%, >40%, >50%, >60%, >70%, >80% or >90% more preferably at least about >95%, >99% or >99.5%, to the PERK sequence in figure 1 (or a partial sequence thereof). Modified polypeptides molecules are discussed below. Preferably about: 1, 2, 3, 4, 5, 6 to 10, 10 to 25, 26 to 50 or 51 to 100, or 101 to 250 nucleotides or amino acids are modified.

Nucleic Acid Molecules and Polypeptides Similar to PERK1

Those skilled in the art will recognize that the nucleic acid molecule sequences in figure 1 are not the only sequences which may be used to provide increased PERK activity in plants. The genetic code is degenerate so other nucleic acid molecules which encode a polypeptide identical to an amino acid sequence in figure 1 may also be used. The sequence of the other nucleic acid molecules of this invention may also be varied without changing the polypeptide encoded by the sequence. Consequently, the nucleic acid molecule constructs described below and in the

accompanying examples for the preferred nucleic acid molecules, vectors, and transformants of the invention are merely illustrative and are not intended to limit the scope of the invention.

The sequences of the invention can be prepared according to numerous techniques. The invention is not limited to any particular preparation means. For example, the nucleic acid molecules of the invention can be produced by cDNA cloning, genomic cloning, DNA synthesis, polymerase chain reaction (PCR) technology, or a combination of these approaches (Current Protocols in Molecular Biology (F. M. Ausbel et al., 1989).). Sequences may be synthesized using well known methods and equipment, such as automated synthesizers. Nucleic acid molecules may be amplified by the polymerase chain reaction. Polypeptides may, for example, be synthesized or produced recombinantly.

Sequence Identity

The invention includes modified nucleic acid molecules with a sequence identity at least about: >17%, >20%, >30%, >40%, >50%, >60%, >70%, >80% or >90% more preferably at least about >95%, >99% or >99.5%, to a DNA sequence in figure 1 (or a partial sequence thereof). Preferably about 1, 2, 3, 4, 5, 6 to 10, 10 to 25, 26 to 50 or 51 to 100, or 101 to 250 nucleotides or amino acids are modified. Identity is calculated according to methods known in the art. Sequence identity is most preferably assessed by the FASTA 3 program. For example, if a nucleotide sequence (called "Sequence A") has 90% identity to a portion of the nucleotide sequence in Figure 1, then Sequence A will be identical to the referenced portion of the nucleotide sequence in Figure 1, except that Sequence A may include up to 10 point mutations, such as substitutions with other nucleotides, per each 100 nucleotide of the referenced portion of the nucleotide sequence in Figure 1. Nucleotide sequences functionally equivalent to the PERK1 sequence can occur in a variety of forms as described below. Polypeptides having sequence identity may be similarly identified.

The polypeptides encoded by the homologous PERK nucleic acid molecule in other species will have amino acid sequence identity (also known as homology) at least about: >20%, >25%, >28%, >30%, >40% or >50% to an amino acid sequence shown in figure 1 (or a partial sequence thereof). Some plants species may have polypeptides with a sequence identity of at least about: >60%, >70%, >80% or >90%, more preferably at least about: >95%, >99% or

>99.5% to all or part of an amino acid sequence in figure 1 (or a partial sequence thereof). Identity is calculated according to methods known in the art. Sequence identity is most preferably assessed by FASTA 3 program. Preferably about: 1, 2, 3, 4, 5, 6 to 10, 10 to 25, 26 to 50 or 51 to 100, or 101 to 250 nucleotides or amino acids are modified.

The invention includes nucleic acid molecules with mutations that cause an amino acid change in a portion of the polypeptide not involved in providing PERK activity or an amino acid change in a portion of the polypeptide involved in providing PERK activity so that the mutation increases or decreases the activity of the polypeptide.

Hybridization

Other functional equivalent forms of the PERK nucleic acid molecules encoding nucleic acids can be isolated using conventional DNA-DNA or DNA-RNA hybridization techniques. These nucleic acid molecules and the PERK sequences can be modified without significantly affecting their activity.

The present invention also includes nucleic acid molecules that hybridize to one or more of the sequences in figure 1 (or a partial sequence thereof) or their complementary sequences, and that encode peptides or polypeptides exhibiting substantially equivalent activity as that of an PERK polypeptide produced by the DNA in figure 1 or its variants. Such nucleic acid molecules preferably hybridize to PERK1 under low, moderate (intermediate), or high stringency conditions (see Sambrook et al. (Most recent edition) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). Preferable hybridization conditions are low stringency conditions of at least about: 1XSSC, 0.1% SDS at 50°C or high stringency conditions of at least about: 0.1XSSC, 0.1% SDS at 65°C.

The present invention also includes nucleic acid molecules from any source, whether modified or not, that hybridize to genomic DNA, cDNA, or synthetic DNA molecules that encode the amino acid sequence of a PERK polypeptide, or genetically degenerate forms, under salt and temperature conditions equivalent to those described in this application, and that code for a peptide, or polypeptide that has PERK activity. Preferably the polypeptide has the same or similar activity as that of a PERK polypeptide. A nucleic acid molecule described above is

considered to be functionally equivalent to a PERK nucleic acid molecule (and thereby having PERK activity) of the present invention if the polypeptide produced by the nucleic acid molecule displays the following characteristics: The defining feature of PERK polypeptides is the presence of a proline-rich domain, followed by a transmembrane domain, followed by a kinase domain. When tested, the kinase domain has serine/threonine kinase activity.

The invention also includes nucleic acid molecules and polypeptides having sequence similarity taking into account conservative amino acid substitutions. Sequence similarity (and preferred percentages) are discussed below.

Modifications to Nucleic Acid Molecule or Polypeptide Sequence

Changes in the nucleotide sequence which result in production of a chemically equivalent or chemically similar amino acid sequences are included within the scope of the invention. Variants of the polypeptides of the invention may occur naturally, for example, by mutation, or may be made, for example, with polypeptide engineering techniques such as site directed mutagenesis, which are well known in the art for substitution of amino acids. For example, a hydrophobic residue, such as glycine can be substituted for another hydrophobic residue such as alanine. An alanine residue may be substituted with a more hydrophobic residue such as leucine, valine or isoleucine. A negatively charged amino acid such as aspartic acid may be substituted for glutamic acid. A positively charged amino acid such as lysine may be substituted for another positively charged amino acid such as arginine.

Therefore, the invention includes polypeptides having conservative changes or substitutions in amino acid sequences. Conservative substitutions insert one or more amino acids which have similar chemical properties as the replaced amino acids. The invention includes sequences where conservative substitutions are made that do not destroy PERK activity. The preferred percentage of sequence similarity for sequences of the invention includes sequences having at least about: 50% similarity to PERK1. The similarity may also be at least about: 60% similarity, 75% similarity, 80% similarity, 90% similarity, 95% similarity, 97% similarity, 98% similarity, 99% similarity, or more preferably at least about 99.5% similarity, wherein the polypeptide has PERK activity. The invention also includes nucleic acid molecules encoding polypeptides, with the polypeptides having at least about: 50% similarity to PERK1. The

similarity may also be at least about: 60% similarity, 75% similarity, 80% similarity, 90% similarity, 95% similarity, 97% similarity, 98% similarity, 99% similarity, or more preferably at least about 99.5% similarity, wherein the polypeptide has PERK activity, to an amino acid sequence in figure 1 (or a partial sequence thereof) considering conservative amino acid changes, wherein the polypeptide has PERK activity. Sequence similarity is preferably calculated number of similar amino acids in a multiple alignment expressed as a percentage of the shorter of the two sequences in the alignment. The multiple alignment is preferably constructed using the FASTA 3 program, using the following parameter settings: gap penalty (open) = -12(protein) – 16 (DNA), gap penalty (extension) = -2 (protein) –4 (DNA) , protein weight matrix = BLOSUM 62. (The reference for FASTA 3 is W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444- 2448,and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA" Methods in Enzymology 183:63-98).

Polypeptides comprising one or more d-amino acids are contemplated within the invention. Also contemplated are polypeptides where one or more amino acids are acetylated at the N-terminus. Those of skill in the art recognize that a variety of techniques are available for constructing polypeptide mimetics with the same or similar desired PERK activity as the corresponding polypeptide compound of the invention but with more favorable activity than the polypeptide with respect to solubility, stability, and/or susceptibility to hydrolysis and proteolysis. See, for example, Morgan and Gainor, Ann. Rep. Med. Chem., 24:243-252 (1989). Examples of polypeptide mimetics are described in U.S. Patent Nos. 5,643,873. Other patents describing how to make and use mimetics include, for example in, 5,786,322, 5,767,075, 5,763,571, 5,753,226, 5,683,983, 5,677,280, 5,672,584, 5,668,110, 5,654,276, 5,643,873. Mimetics of the polypeptides of the invention may also be made according to other techniques known in the art. For example, by treating a polypeptide of the invention with an agent that chemically alters a side group by converting a hydrogen group to another group such as a hydroxy or amino group. Mimetics preferably include sequences that are either entirely made of amino acids or sequences that are hybrids including amino acids and modified amino acids or other organic molecules.

The invention also includes hybrid nucleic acid molecules and polypeptides, for example where a nucleotide sequence from one species of plant is combined with a nucleotide sequence from another sequence of plant, mammal or yeast to produce a fusion polypeptide. The invention includes a fusion protein having at least two components, wherein a first component of the fusion protein comprises a polypeptide of the invention, preferably a full length PERK polypeptide. The second component of the fusion protein preferably comprises a tag, for example GST, an epitope tag or an enzyme. The fusion protein may comprise lacZ.

The invention also includes polypeptide fragments of the polypeptides of the invention which may be used to confer PERK activity if the fragments retain activity. The invention also includes polypeptides fragments of the polypeptides of the invention which may be used as a research tool to characterize the polypeptide or its activity. Such polypeptides preferably consist of at least 5 amino acids. In preferred embodiments, they may consist of 6 to 10, 11 to 15, 16 to 25, 26 to 50, 51 to 75,76 to 100 or 101 to 250 amino acids of the polypeptides of the invention (or longer amino acid sequences). The fragments preferably have PERK activity. Fragments may include sequences with one or more amino acids removed, for example, C-terminus amino acids in a PERK sequence.

The invention also includes a composition comprising all or part of an isolated PERK nucleic acid molecule (preferably PERK1) of the invention and a carrier, preferably in a composition for plant transformation. The invention also includes a composition comprising an isolated PERK polypeptide (preferably PERK1) and a carrier, preferably for studying polypeptide activity.

Recombinant Nucleic Acid Molecules

The invention also includes recombinant nucleic acid molecules comprising a nucleic acid molecule of the invention and a promoter sequence, operatively linked so that the promoter enhances transcription of the nucleic acid molecule in a host cell (the nucleic acid molecules of the invention may be used in an isolated native gene or a chimeric gene (for example, where a nucleic acid molecule coding region is connected to one or more heterologous sequences to form a gene). The promoter sequence is preferably a constitutive promoter sequence or an inducible promoter sequence, operatively linked so that the promoter enhances transcription of the DNA

molecule in a host cell. The promoter may be of a type not naturally associated with the cell such as a super promoter, a 35S cauliflower mosaic virus promoter, a chemical inducible promoter, a copper-inducible promoter, a steroid-inducible promoter and a tissue specific promoter .

A recombinant nucleic acid molecule for conferring PERK activity may also contain suitable transcriptional or translational regulatory elements. Suitable regulatory elements may be derived from a variety of sources, and they may be readily selected by one with ordinary skill in the art. Examples of regulatory elements include: a transcriptional promoter and enhancer or RNA polymerase binding sequence, a ribosomal binding sequence, including a translation initiation signal. Additionally, depending on the vector employed, other genetic elements, such as selectable markers, may be incorporated into the recombinant molecule. Markers facilitate the selection of a transformed host cell. Such markers include genes associated with temperature sensitivity, drug resistance, or enzymes associated with phenotypic characteristics of the host organisms.

Nucleic acid molecule expression levels are controlled with a transcription initiation region that regulates transcription of the nucleic acid molecule or nucleic acid molecule fragment of interest in a plant, bacterial or yeast cell. The transcription initiation region may be part of the construct or the expression vector. The transcription initiation domain or promoter includes an RNA polymerase binding site and an mRNA initiation site. Other regulatory regions that may be used include an enhancer domain and a termination region. The regulatory elements described above may be from animal, plant, yeast, bacterial, fungal, viral or other sources, including synthetically produced elements and mutated elements.

Methods of modifying DNA and polypeptides, preparing recombinant nucleic acid molecules and vectors, transformation of cells, expression of polypeptides are known in the art. For guidance, one may consult the following US patent nos. 5,840,537, 5,850,025, 5,858,719, 5,710,018, 5,792,851, 5,851,788, 5,759,788, 5,840,530, 5,789,202, 5,871,983, 5,821,096, 5,876,991, 5,422,108, 5,612,191, 5,804,693, 5,847,258, 5,880,328, 5,767,369, 5,756,684, 5,750,652, 5,824,864, 5,763,211, 5,767,375, 5,750,848, 5,859,337, 5,563,246, 5,346,815, and WO9713843. Many of these patents also provide guidance with respect to experimental assays, probes and antibodies, methods, transformation of host cells and regeneration of plants, which

are described below. These patents, like all other patents, publications (such as articles and database publications) in this application, are incorporated by reference in their entirety.

Host Cells Including a PERK Nucleic Acid Molecule

In a preferred embodiment of the invention, a plant or yeastcell is transformed with a nucleic acid molecule of the invention or a fragment of a nucleic acid molecule and inserted in a vector.

Another embodiment of the invention relates to a method of transforming a host cell with a nucleic acid molecule of the invention or a fragment of a nucleic acid molecule, inserted in a vector. The invention also includes a vector comprising a nucleic acid molecule of the invention. The nucleic acid molecules can be cloned into a variety of vectors by means that are well known in the art. The recombinant nucleic acid molecule may be inserted at a site in the vector created by restriction enzymes. A number of suitable vectors may be used, including cosmids, plasmids, bacteriophage, baculoviruses and viruses. Suitable vectors are capable of reproducing themselves and transforming a host cell. The invention also relates to a method of expressing polypeptides in the host cells. A nucleic acid molecule of the invention may be used to transform virtually any type of plant, including both monocots and dicots. The expression host may be any cell capable of expressing PERK, such as a cell selected from the group consisting of a seed (where appropriate), plant cell, bacterium, yeast, fungus, protozoa, algae, animal and animal cell.

Levels of nucleic acid molecule expression may be controlled with nucleic acid molecules or nucleic acid molecule fragments that code for anti-sense RNA inserted in the vectors described above.

Agrobacterium tumefaciens-mediated transformation, particle-bombardment-mediated transformation, direct uptake, microinjection, coprecipitation and electroporation-mediated nucleic acid molecule transfer are useful to transfer a PERK nucleic acid molecule into seeds (where appropriate) or host cells, preferably plant cells, depending upon the plant species. The invention also includes a method for constructing a host cell capable of expressing a nucleic acid molecule of the invention, the method comprising introducing into said host cell a vector of the

invention. The genome of the host cell may or may not also include a functional PERK gene. The invention also includes a method for expressing a PERK polypeptide in the host cell or a plant, plant part, seed or plant cell of the invention, the method comprising culturing the host cell under conditions suitable for gene expression. The method preferably also includes recovering the expressed polypeptide from the culture.

The invention includes the host cell comprising the recombinant nucleic acid molecule and vector as well as progeny of the cell. Preferred host cells are fungal cells, yeast cells, bacterial cells, mammalian cells, bird cells, reptile cells, amphibious cells, microorganism cells and plant cells. Host cells may be cultured in conventional nutrient media. The media may be modified as appropriate for inducing promoters, amplifying genes or selecting transformants. The culture conditions, such as temperature, composition and pH will be apparent. After transformation, transformants may be identified on the basis of a selectable phenotype. A selectable phenotype can be conferred by a selectable marker in the vector.

Transgenic Plants and Seeds

Plant cells are useful to produce tissue cultures, seeds or whole plants. The invention includes a plant, plant part, seed, or progeny thereof including a host cell transformed with a PERK nucleic acid molecule. The plant part is preferably a leaf, a stem, a flower, a root, a seed or a tuber.

The invention includes a transformed (transgenic) plant having increased PERK activity, the transformed plant containing a nucleic acid molecule sequence encoding for polypeptide activity and the nucleic acid molecule sequence having been introduced into the plant by transformation under conditions whereby the transformed plant expresses a PERK polypeptide in active form.

The methods and reagents for producing mature plants from cells are known in the art. The invention includes a method of producing a genetically transformed plant which expresses PERK polypeptide by regenerating a genetically transformed plant from the plant cell, seed or plant part of the invention. The invention also includes the transgenic plant produced according to the method. Alternatively, a plant may be transformed with a vector of the invention.

The invention also includes a method of preparing a plant with increased PERK activity, the method comprising transforming the plant with a nucleic acid molecule which encodes a PERK polypeptide or a polypeptide encoding a PERK polypeptide capable of increasing PERK activity in a cell, and recovering the transformed plant with increased PERK activity. The invention also includes a method of preparing a plant with increased PERK activity, the method comprising transforming a plant cell with a nucleic acid molecule which encodes a PERK polypeptide capable of increasing PERK activity in a cell.

Overexpression of PERK leads to an improved ability of the transgenic plants to resist wounding or pathogen damage.

The plants whose cells may be transformed with a nucleic acid molecule of this invention and used to produce transgenic plants include, but are not limited to the

Target plants: *Brassica napus, Brassica rapa, Brassica juncea, Brassica oleracea*, or from the family Brassicaecae, Arabidopsis, potato, tomato, tobacco, cotton, carrot, petunia, sunflower, strawberries, spinach, lettuce, rice, soybean, corn, wheat, rye, barley, sorgum and alfalfa. Cereal plants including rye, barley and wheat may also be transformed with a PERK polypeptide, preferably PERK1.

In a preferred embodiment of the invention, plant tissue cells or cultures which demonstrate PERK activity (or increased PERK activity compared to wild type) are selected and plants are regenerated from these cultures. Methods of regeneration will be apparent to those skilled in the art (see Examples below, also). These plants may be reproduced, for example by cross pollination with a plant that does not have PERK activity. If the plants are self-pollinated, homozygous progeny may be identified from the seeds of these plants, for example, using genetic markers. Seeds obtained from the mature plants resulting from these crossings may be planted, grown to sexual maturity and cross-pollinated or self-pollinated.

The nucleic acid molecule is also incorporated in some plant species by breeding methods such as back crossing to create plants homozygous for the PERK nucleic acid molecule.

A plant line homozygous for the PERK nucleic acid molecule may be used as either a male or female parent in a cross with a plant line lacking the PERK nucleic acid molecule to

produce a hybrid plant line which is uniformly heterozygous for the nucleic acid molecule. Crosses between plant lines homozygous for the PERK nucleic acid molecule are used to generate hybrid seed homozygous for the resistance nucleic acid molecule.

Fragments/Probes

Preferable fragments include 10 to 50, 50 to 100, 100 to 250, 250 to 500, 500 to 1000, 1000 to 1500, or 1500 or more nucleotides of a nucleic acid molecule of the invention. A fragment may be generated by removing a single nucleotide from a sequence in figure 1 (or a partial sequence thereof). Fragments may or may not encode a polypeptide having PERK activity.

The nucleic acid molecules of the invention (including a fragment of a sequence in figure 1 (or a partial sequence thereof) can be used as probes to detect nucleic acid molecules according to techniques known in the art (for example, see US patent nos. 5,792,851 and 5,851,788). The probes may be used to detect nucleic acid molecules that encode polypeptides similar to the polypeptides of the invention. For example, a probe having at least about 10 bases will hybridize to similar sequences under stringent hybridization conditions (Sambrook et al. 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor).

Kits

The invention also includes a kit for conferring increased PERK activity to a plant or a host cell including a nucleic acid molecule of the invention (preferably in a composition of the invention) and preferably reagents for transforming the plant or host cell.

The invention also includes a kit for detecting the presence of a PERK or a PERK nucleic acid molecule, comprising at least one probe of the invention. Kits may be prepared according to known techniques, for example, see patent nos. 5,851,788 and 5,750,653.

Antibodies

The invention includes an isolated antibody immunoreactive with a polypeptide of the invention. The antibody may be labeled with a detectable marker or unlabeled. The antibody is preferably a monoclonal antibody or a polyclonal antibody. PERK antibodies can be employed to

screen organisms containing PERK polypeptides. The antibodies are also valuable for immunopurification of polypeptides from crude extracts.

Examples of the preparation and use of antibodies are provided in US Patent Nos. 5,792,851 and 5,759,788. For other examples of methods of the preparation and uses of monoclonal antibodies, see US Patent Nos. 5,688,681, 5,688,657, 5,683,693, 5,667,781, 5,665,356, 5,591,628, 5,510,241, 5,503,987, 5,501,988, 5,500,345 and 5,496,705. Examples of the preparation and uses of polyclonal antibodies are disclosed in US Patent Nos. 5,512,282, 4,828,985, 5,225,331 and 5,124,147.

The invention also includes methods of using the antibodies. For example, the invention includes a method for detecting the presence of a PERK polypeptide, by: a) contacting a sample containing one or more polypeptides with an antibody of the invention under conditions suitable for the binding of the antibody to polypeptides with which it is specifically reactive; b) separating unbound polypeptides from the antibody; and c) detecting antibody which remains bound to one or more of the polypeptides in the sample.

Research Tool

Cell cultures, seeds, plants and plant parts transformed with a nucleic acid molecule of the invention are useful as research tools. For example, one may obtain a plant cell (or a cell line, such as an immortalized cell culture or a primary cell culture) that does not express PERK1, insert a PERK1 nucleic acid molecule in the cell, and assess the level of PERK1 expression and activity.

The PERK nucleic acid molecules and polypeptides are also useful in assays. Assays are useful for identification and development of compounds to inhibit and/or enhance polypeptide function directly.

Suitable assays may be adapted from, for example, US patent no. 5,851,788

Using Exogenous Agents in Combination with a Vector

The nucleic acid molecules of the invention may be used with other nucleic acid molecules that relate to plant protection, for example, extensin nucleic acid molecules. Host cells or plants may be transformed with these nucleic acid molecules.

PERK1 POLYPEPTIDE

A PERK1 antibody is an invaluable tool in studies to address the expression, localization, regulation and potential function of PERK1. The coding sequence for the region containing the extracellular domain of PERK1 as well as the coding sequence representing the entire protein were cloned into the pTrcHis expression plasmid for production of His-tagged fusion proteins in *E. coli* and purification by affinity chromatography on a Talon resin column.

Figure 8A represents a Western blot performed on the bacterially expressed extracellular domain fusion protein using a T7 monoclonal antibody which recognizes the T7 epitope on the bacterially expressed fusion protein . There was a marked induction of the fusion protein in the presence of IPTG (Isopropyl-beta-d-Thiogalactopyranoside) (Figure 8A; lane 2), however recovery of the purified protein (Lane 4; bold-face diamond) was not very efficient. The presence of the fusion protein in the insoluble fraction (Lane 3) shows that the extracellular domain is likely insoluble and could be purified from inclusion bodies. This method relies on the bacterial expression system generating large amounts the insoluble fusion protein which form inclusion bodies in the bacterial host. Figure 8B is a Western blot to confirm that the extracellular domain fusion protein is targeted to inclusion bodies. Induction of the fusion protein was efficient (Figure 8B; lane 2), and despite its presence in the supernatant (Lane 3) and in the denatured insoluble samples (Lane 5), the fusion protein is predominantly in the insoluble fraction (IB-inclusion body) (Lane 4; open-face diamond) Lanes 6-9 represent the pTrcHisC negative control indicating that the signals obtained for the expression of the extracellular domain protein are specific to the fusion protein. The extracellular domain fusion protein can be used as an antigen and injected into rabbits in order to generate a polyclonal antibody against the extracellular domain of PERK1. This fusion protein can either be injected directly as an inclusion body in the presence of an adjuvant, or alternatively, the protein purified from the inclusion bodies can be separated on an SDS-Polyacrylamide gel, and the region of the gel at which the protein of interest migrates can be excised and used for injection. It should be noted that the bacterially expressed fusion protein is migrating at a higher molecular weight than predicted from the amino acid sequence, however this is not entirely uncommon and could in fact

be due to the high percentage of proline residues (40%) comprising the extracellular domain of PERK1.

Similar experiments were conducted for the fusion protein corresponding to the entire coding region of PERK1. Figure 9 is a Western blot conducted on the bacterially expressed PERK1 protein to determine whether the full length protein is insoluble and forms inclusion bodies. Although the induction of the full length protein was not as pronounced (Figure 9; lane 2) it is quite evident that the fusion protein is predominantly localized in the insoluble fraction (i.e. inclusion body) (Lane 4; bold-face square) Lanes 6-9 represent the parental plasmid as mentioned above. We are currently modifying several variables of the induction protocol in an attempt to maximize the yield of full length PERK1 protein The PERK1 fusion protein can also be used as an antigen to raise antibody against full length PERK1 as previously described.

We show that PERK1 encodes a protein with kinase activity as its sequence predicts. The bacterially expressed catalytic domain fusion protein of PERK1 is tested for kinase activity. Furthermore, to ensure that the phosphorylation of the fusion protein was not a result of bacterial kinase activity, a mutated catalytic domain was also generated by site directed mutagenesis which introduced a single base pair substitution of a lysine residue to a glutamic acid residue (K→E). This mutation modifies the essential invariant lysine of subdomain II required for phospho-transfer and renders the kinase inactive. Both the wild-type and the mutated catalytic domains of PERK1 were cloned into the pMAL-c expression system, induced for protein production in the presence of IPTG and purified by affinity chromatography on MBP amylose resin. Figure 10A is a Western blot to confirm the induction and purification of both the wild-type (bold-face star) and mutated (bold-face arrow) catalytic domain fusion proteins using an anti-MBP antibody. The wild-type fusion protein appears to be toxic in bacteria which compromises its inducibility and purification (Figure 10A; lanes 3-4). In addition, the wild-type fusion protein migrates at a slightly larger molecular weight than does the mutated protein. The mutated fusion protein is induced and purified more efficiently, perhaps due to the fact that it is no longer kinase active (Figure 10A ; Lanes 6-7). Purified pMAL-c (Lane 1; bold-face circle) was used as a negative control for this experiment.

Figure 10B represents a kinase assay performed on affinity purified wild-type and mutated fusion proteins incubated in the presence of γ-32PdATP. Detection of a phosphoprotein only in Lane 1 provides direct biochemical evidence that the wild-type catalytic domain (bold-face star) of PERK1 encodes a functional protein kinase that is capable of autophosphorylation (Lane 1) and that the mutation successfully abolished kinase activity (Lane 2).

Since PERK1 is known to encode a protein with kinase activity phosphoamino acid analysis was performed to determine the amino acid specificity of its autophosphorylating activity. The results demonstrate that PERK1 is phosphorylated on serine/threonine residues and is therefore a serine/threonine receptor kinase.

Transgenic Work

An important focus of agricultural biotechnology research is in devising new strategies to combat crop losses due to plant diseases and pests. Advances in the development and improvement of plant transformation techniques has opened up new avenues for generating crops with enhanced resistance against disease and insect attack (i.e. mechanical wounding). Plants can now be engineered to express pathogen derived compounds that disrupt the infection process or alternatively a more desirable and perhaps effective approach is to enhance a plant's endogenous defense mechanisms. Furthermore, the downstream defense mechanisms are fairly well understood, therefore manipulation of some potential upstream signals that control the battery of defense proteins may be a promising strategy to engineer plants with enhanced and broad-spectrum resistance against both insect injury and pathogen attack. PERK1 functions as one such gene.

One approach to show the role of a particular gene in the regulation of gene expression in response to pathogen attack or mechanical wounding has been to generate transgenic plants constitutively expressing the respective cDNA in the sense [1-2] or antisense [3] orientations and examining the effects of wounding and pathogen attack on known downstream target genes involved in these processes. Generating plants that overexpress the antisense PERK1 transcript shows the role of PERK1 in mediating a plant's defense response to both wounding and pathogen attack. By mechanically injuring, inoculating with pathogens (i.e. *Sclerotinia sclerotiorum, Cylindrosporium concertricum, Phoma lingam*) or treating with chemical elicitors

and looking at the levels of downstream genes directly involved in these various processes, we directly implicate PERK1 in these pathways. For example, the expression of the functional PERK1 protein is abolished and induction of downstream genes is reduced, so PERK1 is an important upstream component of the pathway. PERK1 offers protection against wounding and pathogen attack, so plants overexpressing PERK1 in the sense orientation exhibit an accumulation of downstream target transcripts involved in these responses and ultimately enhanced resistance.

We score for a phenotype such as enhanced survival of plants overexpressing PERK1 compared to the enhanced susceptibility of plants overexpressing antisense PERK1 transcripts in response to pathogen treatment.

(References 1. Tang, X. et al. (1999). Overexpression of *Pto* activates defense responses and confers broad resistance. Plant Cell 11, 15-29; 2. Cao, H. et al. (1998). Generation of broad-spectrum disease resistance by overexpression of an essential regulatory gene in systemic acquired resistance. Proc. Natl. Acad. Sci. 95, 6531-6536; 3. Royo, J. et al. (1999). Antisense-mediated depletion of a potato lipoxygenase reduces wound induction of proteinase inhibitors and increases weight gain of insect pests. Proc. Natl. Acad. Sci. 96, 1146-1151.)

Generating Transgenic PERK1 *Brassica napus* Plants

In order to generate transgenic *Brassica napus* plants expressing altered levels of PERK1, the full length PERK1 cDNA is cloned into a plant transformation vector in the sense orientation downstream of the constitutively active 35S promoter from the cauliflower mosaic virus (35S CaMV). Subsequently, this expression construct is introduced into an *Agrobacterium tumefaciens* strain in order to transform *B. napus* plants via *Agrobacterium* mediated transformation. Plant transformants containing the integrated PERK1 cDNA are selected and tested for the presence of the transgene.

Upon generating transgenic plants constitutively expressing the PERK1 cDNA in the sense orientation, we examine the effects of wounding and pathogen attack on known downstream genes involved in these processes to confirm the involvement of PERK1 in these pathways. PERK1 is an upstream component of these pathways. It is involved in mediating a

plant's defense response to both wounding and pathogen attack , and transgenic plants overexpressing the PERK1 cDNA in the sense orientation exhibit an increase in the expression of downstream target genes of these pathways. Furthermore, these plants exhibit an enhanced survival relative to a wild type plant in response to wounding and pathogen attack.

We investigate the role of PERK1 as a signaling molecule in response to wounding and pathogen attack. Induction of PERK1 with respect to changes in levels of phosphorylation shows the role of PERK1 in these processes. The PERK1 protein is immunoprecipitated using the anti-PERK1 antibody from total protein extracts prepared from *B. napus* tissue after treatment with a specific stimulus and immunoblotted using a commercially available anti-phospho (serine or threonine) antibody. This approach provides quantitative results for the levels of PERK1 phosphorylation in response to various stimuli.

EXAMPLES

Isolation and Sequence Analyses of PERK1 cDNA

In order to isolate novel receptor-like protein kinases in *B. napus* a combination of degenerate oligonucleotide primers designed against conserved kinase subdomains I and VII (Hanks and Quinn, 1991) were used to amplify mass excised phagemid DNA from a newly constructed lambda-pistil cDNA library. The cDNAs encoding products of the expected length (~420-450 bp) were cloned and the deduced partial sequences were analysed against several databases in order to determine which clone represented a novel kinase. One of several candidates, showed the highest degree of sequence similarity to the cytoplasmic domain of known plant receptor protein kinases, and was therefore used to screen the amplified lambda-pistil cDNA library. Several positive clones obtained from the library screen were completely sequenced and a partial 1512 bp consensus sequence was generated to represent the PERK1 cDNA isolated from the library screen. Although this partial PERK1 cDNA had an open reading frame, it did not encode a full length transcript, therefore the 5' end was completed by 5' RACE (see Methods).

The deduced amino acid sequence of PERK1 is shown in Figure 1A and a schematic representation of its hydropathy plot is shown in Figure 1B. The full length cDNA sequence is

2189 bp and consists of one large open reading frame of 1944 bp encoding a predicted protein of 648 amino acids with an estimated molecular mass of 69 kDa (Figure 1). The first methionine of this open reading frame is preceded by two in frame stop codons, TAA and TGA at positions -48 to -45 and -23 to -21 respectively. In addition, there is also an AGAA sequence at position -9 to -6 (Figure 1) which is a favourable site for translation initiation in all eukaryotes (Lutcke et al., 1987).

PERK1 encodes a receptor-like kinase possessing an extracellular domain, a single membrane spanning domain and an intracellular kinase domain (Figure 1B) with four potential N-linked glycosylation sites (Asn-X-Ser/Thr) found throughout the sequence (Figure 1A) (Weinstein et al., 1982). The predicted polypeptide sequence was analyzed using the PSORT database and determined to be a Type Ib intergral membrane protein with a hydrophilic amino terminal domain exposed on the exterior of the membrane but whose coding sequence does not indicate a cleavable signal sequence preceding this domain. Singer (1990) proposes that despite the lack of a signal peptide, Type Ib integral membrane proteins are inserted into the membrane via the usual ER-translocator protein machinery with some slight modifications. The extracellular domain of this protein consists of 137 amino acids (Figure 1A) rich in proline and sharing sequence similarity with extensins, a family of hydroxyproline-rich glycoproteins (HRGPs) that constitute a major protein component of higher plant cell walls (Showalter, 1993). Extensin proteins have two proposed functions in plants, one which contributes to the structural support of the cell wall by forming glycoprotein networks and the other which involves plant defense; helping to protect the plant against mechanical wounding or pathogen attack (Wilson and Fry, 1986; Showlater, 1993). A distinctive characteristic prevalent among dicot extensins is the repetitive Ser-$(Pro)_4$ pentapeptide consensus motif (Showalter, 1993). A unique feature of PERK1's extracellular domain is the predominance of a slightly modified Ser-$(Pro)_{2-3}$ motif in addition to the presence of one signature pentapeptide block (Figure 1A). In order to investigate the phylogenetic status of PERK1, sixty four deduced amino sequences corresponding to the extracellular and transmembrane domains of extensin, proline rich and other cell wall proteins were retrieved from Genbank and used to construct a phylogenetic tree (Clustal X). Results from the phylogenetic analysis indicated that PERK1 is most similar to a subset of extensin proteins as shown by the sequence homology restricted predominantly to the serine/proline rich regions of

the protein (data not shown). Extensins and proline rich proteins comprise two major classes of cell wall proteins and are essential for maintaining the proper architecture of a plant cell wall as well as important in helping protect plant cells against wounding and pathogen invasion. These protein families have been the focus of many research efforts and members belonging to these protein classes have been isolated in a wide range of plant systems. A certain degree of sequence identity or conservation is retained among these proteins isolated from different plants. Given that PERK1 shows homology in its extracellular domain to the extensin family of cell wall proteins and is rich in proline residues, homologues of PERK1 exist in many other plant systems.

The protein also contains two other domains of note. Hydropathy analysis (Kyte and Doolittle, 1982) of the protein sequence predicted a membrane spanning region of 23 amino acids (Figure 1; residues 138-160) followed by a characteristic stop transfer sequence rich in charged amino acids [Arg-Arg-Arg] required for the proper insertion in the membane (Weinstein et al., 1982).

All known protein kinases display amino acid sequence similarity in their catalytic domains which are comprised of eleven subdomains containing some invariant residues important for catalysis (Hanks and Quinn, 1991). The overall features of this organization are identified in the catalytic domain of the PERK1 protein in that all of the absolutely conserved amino acids as well as the highly conserved amino acid groups are present (Figure 1A). The sequences of DIKASN in subdomain VI and GTFGYLAPE in subdomain VIII (Hanks and Quinn, 1991) are strong indicators that PERK1 may possess serine/threonine rather than tyrosine substrate specificity (Figure 1A).

PERK1 is a Single Copy Gene and Ubiquitously Expressed in *B. napus* Tissue

As shown in Figure 2, Southern blot analysis was performed under conditions of varying stringency using *B. napus* genomic DNA digested with several restriction enzymes in order to determine copy number of PERK1 in the Brassica genome. Based on known restriction sites within the cDNA and identical hybridization patterns obtained for low and high stringency conditions (Figure 2), PERK1 appears to exist as a single copy gene and is not a member of a multigene family.

In order to determine whether PERK1 is expressed in plant tissues, RNA gel blot analysis was performed using poly(A)$^+$ mRNA isolated from a variety of *B. napus* tissues as shown in Figure 3. A partial 1.5 kb PERK1 cDNA probe used in this experiment detected a transcript of 2.2 kb (Figure 3A) which is consistent with the size of the full length PERK1 cDNA. The presence of a higher molecular weight transcript of ~ 4.4 kb in root (Figure 3A) may represent an alternatively spliced transcript of PERK1 generated during mRNA processing. This is reminiscent of the alternative transcript detected for the carrot extensin gene which as reported by Chen and Varner (1985) is predominantly expressed in root tissue.

The 2.2 kb PERK1 transcript was most abundant in *B. napus* root, stem and petal tissue (Figure 3A). Levels of PERK1 mRNA were also detected in leaf and pistil tissue albeit at much lower levels and transcripts appear to be absent in the anther tissue (Figure 3A). As an internal control the blot was reprobed with a cyclophilin EST cDNA to ensure even loading of the poly(A)$^+$ mRNA. A 700 bp cyclophilin transcript detected with relatively the same intensity in all tissues with the exception of anther indicates that equal amounts of mRNA was used (Figure 3B). In addition, the ethidium bromide stained gel shows relatively even loading (Figure 3C). The striking difference in the intensity of the cyclophilin signal in anther tissue (Figure 3B) is a common problem associated with the desiccate nature of the tissue, and could be attributed to the low abundance of cyclophilin mRNA present in the anther tissue as reported by Gasser et al. (1990).

Changes in PERK1 mRNA in Response to Wounding and Chemical Elicitors

In order to examine whether PERK1 expression could be influenced by external stimuli, leaf and stem tissue of *B. napus* plants were wounded and the abundance of PERK1 mRNA was determined by standard Northern blot analysis using the full length PERK1 cDNA as a probe (see Methods). Figure 4 shows changes in the steady-state levels of PERK1 mRNA accumulation following injury. PERK1 transcripts in wounded leaf tissue began to accumulate 5 min after wounding, reaching maximal levels within 15 min post injury represented by an 8.2 fold induction. A 4.5 fold increase in PERK1 mRNA levels was detected 45 min following treatment declining towards basal levels by 1 hr (Figure 4A).

A similar profile of PERK1 mRNA steady state levels was obtained for wounded stem tissue (Figure 4B). An accumulation of PERK1 mRNA in stem is evident 15 min following wounding which represents a 3.6 fold induction of this gene. Maximum steady state levels of PERK1 mRNA in stem was achieved 45 min after injury corresponding to a 4.5 fold induction. Therefore, the overall kinetics of PERK1 mRNA accumulation in both tissues after mechanical wounding is clearly a very rapid response (Figure 4).

Defense mechanisms deployed by plants in response to wounding or pathogen attack have been shown to be induced by certain plant derived chemicals such as methyl jasmonate (MeJA) and salicylic acid (SA). In order to examine changes in the levels of PERK1 mRNA abundance in response to exogenous application of MeJA, *B. napus* plants were thoroughly sprayed with a 5 micromolar MeJA solution. Leaf and stem tissue was subsequently harvested at various times and the steady state levels of PERK1 mRNA were analysed. Figure 5A shows the RNA gel blot and corrected PERK1 mRNA profile for treated leaf tissue during which no significant accumulation of PERK1 mRNA was detected. In response to MeJA, levels of PERK1 transcript in leaf tissue were very weak resembling basal levels in untreated tissue (Figure 3A). Exogenous application of MeJA to stem tissue had no effect on the accumulation of PERK1 mRNA as shown by the corrected profile in which the fold induction of PERK1 did not deviate substantially from the untreated control (0 hr) (Figure 5B). Furthermore, no increase in the steady state levels of PERK1 mRNA was detected in the appropriate control treatment (0.1% [$^{v}/_{v}$] ethanol, solvent control for MeJA) at time 0 hr.

Many genes isolated to date that are induced by a pathogenic stimulus can be at least partially induced by SA (Ward et al., 1991). In order to address the potential role of PERK1 in a plant's defense response against pathogen attack, 4mM SA was used as a chemical elicitor and sprayed onto *B. napus* plants. Figure 6A shows that when SA is exogenously applied to leaf tissue, PERK1 mRNA accumulates 15 min following treatment reaching a maximum 5 fold induction 45 min post-treatment. Steady state levels of PERK1 mRNA in treated stem tissue peaked at 45 min corresponding to an approximate 2 fold induction in response to 4mM SA (Figure 6B).

MATERIALS AND METHODS

Construction of Lambda-Pistil cDNA Library

Pistils were collected from floral buds of Westar and W1 cultivars 1-2 days before anthesis. Total RNA was isolated using the method described by Jones et al. (1985), and enriched for poly(A)$^+$ mRNA by affinity chromatography using pre-packed oligo $(dT)_{25}$-cellulose beads (New England Biolabs, Beverly, MA). Approximately five micrograms of pistil poly(A)$^+$ mRNA was used for the construction of a cDNA library using the ZAP-cDNA® synthesis kit (Stratagene, La Jolla, CA). The information encoded by the poly(A)$^+$ mRNA was reversed transcribed using M-MuLV RT and converted into stable, unidirectional cDNA which was subsequently inserted into a self-replicating Uni-ZAP XR vector, packaged into phage particles in three separate packaging reactions and amplified as described by the manufacturer's procedures (Stratagene, La Jolla, CA). Infection of *Escherichia coli* host strain XLI-Blue yielded a primary library with an average titer of 1.0×10^6 plaque forming units. The primary library was subsequently amplified to obtain an average total of 6.6×10^{10} plaque forming units.

Generation of Novel Receptor-like Protein Kinase Clones

The isolation of novel *Brassica napus* receptor kinases relied upon the newly constructedcDNA library and involved *in vivo* mass excision of the pBluecsript phagemids from the Uni-ZAP XR vectors as outlined by the manufacturer (Stratagene, La Jolla, CA). Following efficient mass excision, phagemid DNA was extracted using a large scale alkaline protocol as described by Sambrook et al. (1989) and subjected to the polymerase chain reaction (PCR) using two separate oligonucleotide combinations, RK1/RK2 and RK1/RK3 (otained from M. Cock, École Normale Supérieure de Lyon, France) specifically designed to prime conserved subdomains of the catalytic domain of receptor protein kinases. RK1 (5- ggiggTTTCggiAT^{T}c$_A$gTiTT$^{A}_{T}{}^{T}_{C}$AA^A_ggg - 3') served as the forward primer and was constructed based upon a conserved amino acid consensus (GGFGIV$^F/_Y$KG) within subdomain I of the catalytic domain. The degeneracy of one reverse primer RK2 (5' - AAiATiC^T_gigCCATiCC^A_gAA$^{A}_{g}{}^{T}_{C}$ - 3') reflects a conserved amino acid consensus (DFGMARIF) of subdomain VII which closely resembles the SRKs in Brassica. The second reverse oligonucleotide RK3 (5' - A^g_AiA$^{g}_{A}{}^{T}_{C}$TTigCiA^A_giCC^A_gAA^A_gTC - 3') was generated based upon conserved amino acids (DFGLAKLL) within subdomain VII prevalent among the

RLKs isolated in Arabidopsis. Phagemid DNA was amplified in a reaction mixture containing 1 microliter of excised phagemid DNA, 10x PCR buffer (100mM Tris-HCl pH8.3, 500mM KCl, 15mM $MgCl_2$), 10mM deoxyribonucleotide triphosphate mixture, 10 micromolar of each oligonucleotide primer and 0.5 microliter Tsg polymerase (BioBasics, Canada). The PCR reaction was heated at 95°C for 2 min and amplified for 35 cycles under the following amplification conditions: 1 min at 95°C for denaturation, 1 min 30 sec at 50°C for primer annealing and 1 min at 72°C for synthesis. A final extension cycle of 10 min at 72°C was also incorporated into the amplification program. All PCR products generated of the expected size (420-450 bp) were gel purified, cloned into the pT7Blue plasmid (Novagen, Madison, WI) and introduced into *Escherichia coli* DH5- alpha. Transformants were tested for the presence of an insert and positive clones were sequenced with universal primers (R-20 and U-19) by an ABI automated sequencer (Model 373 STRETCH DNA; Perkin Elmer Corp., Canada Ltd.) using the dideoxychain-terminating method described by Sanger et al. (1977). Sequence analyses performed using DNAsis® software (Hitachi Software, San Bruno, CA) at the nucleotide and amino acid levels.

Screening of Lambda-Pistil cDNA Library

The original 351 bp PCR product was used to screen the lambda-pistil cDNA library. Approximately $2x10^6$ plaques from the amplifiied library were screened and plated at a density of $1x10^5$ pfu/plate. Duplicate colony lifts were performed according to Sambrook et al. (1989), and prehybridized for 2 hr at 42°C in 50% formamide, 5x Denhardt's solution (1x Denhardt's solution is 0.02% Ficoll, 0.02% DVP, 0.02% BSA), 5x SSC (1x SSC is 0.15M NaCl, 0.015M sodium citrate), 0.1% SDS, 1mM EDTA and 100:g/ml salmon sperm DNA. Filters were subsequently hybridized overnight in the same solution containing the 351 bp PERK1 cDNA radiolabeled by random priming (Feinberg and Vogelstein, 1983) and washed twice with 2x SSC, 0.1% SDS at room temperature for 15 min, followed by two 25 min washes with 0.5x SSC, 0.1% SDS at 55°C. Plaques containing putative positive clones were cored and subjected to several rounds of screening until single isolates representing the PERK1 clone were obtained. Single clone excision to liberate the double stranded pBluescript phagemid was performed on each isolate according to the procedure recommended by the manufacturer (Stratagene, La Jolla,

CA). Phagemid DNA digested with EcoRI/XhoI to release the cloned cDNA was subjected to standard plasmid Southern blot analysis as described by Sambrook et al. (1989) and probed with the radiolabeled 351 bp PERK1 cDNA. The membrane was prehybridized at 42°C in 5x SSPE, 10x Denhardt's solution and 0.5% SDS for 2 hr and hybridized overnight at the same temperature in a buffer containing 50% formamide, 5x SSPE and 0.5% SDS. Washing conditions were performed twice at room temperature for 15 min in 2x SSC, 0.1% SDS followed by several 30 min washes at 55-60°C in 0.1x SSC, 0.1% SDS. An intense hybridization signal would confirm whether phagemids isolated from the library screen contained the cloned cDNA of interest. Several positive clones were sequenced as previously mentioned using both universal and sequence specific primers to generate a consensus sequence representing the PERK1 cDNA clone (1512bp) isolated from the lambda-pistil cDNA library.

Rapid Amplification of cDNA Ends (5'RACE)

The 5' end of the PERK1 cDNA was obtained by the procedure for the rapid amplification of cDNA ends originally described by Frohman et al. (1988) using the 5' RACE System, Version 2.0 kit (Gibco-BRL, Gainthersburg, MD). First strand cDNA was synthesized from approximately 300:g of mixed Westar and W1 pistil total RNA using a gene specific primer GSP1 (5'-TAACCAACAAgAgACA-3') designed to anneal approximately 300 bp from the 5' end of the PERK1 cDNA (1512 bp) isolated from the library screen. Following cDNA synthesis, the first strand product was purified from unincorporated dNTPs and GSP1 using a GLASS MAX® spin cartridge. A homopolymeric tail was added to the 3' end of the cDNA using TdT (terminal deoxynucleotidyl transferase) and dCTP. Tailed cDNA was amplified using a second gene specific primer GSP2 (5'-CCACTCCCAACTTTCAAC -3') desgined to anneal 3' to GSP1 with respect to the cDNA, and an abridged anchor primer (Gibco-BRL, Gainthersburg, MD) which annealed to the homopolymeric tail. PCR amplification was carried out for 35 cycles of denaturation at 94°C for 1 min, primer annealing at 55°C for 1 min and extension at 72°C for 2 min, followed by a final extension cycle for 10 min. A PCR product of the expected size (~1 kb) corresponding to the 5'end of PERK1 was gel purified, cloned into the pT7Blue plasmid (Novagen, Madison, WI) and transformed into *Escherichia coli* DH5- alpha. Confirmation of

the 5'RACE product was obtained by plasmid Southern blot analysis as described above and by sequential primer based sequencing.

Cloning of Full Length PERK1 cDNA

A PCR based approach was used to generate a full length PERK1 cDNA by combining the 5'RACE product cloned into the EcoRV site of pT7Blue with the cDNA isolated from the library screen cloned into the EcoRI/XhoI sites of the pBluescript SK phagemid. A forward primer (5'-ggAAAgCTTgCATgCCTgCAggTCgAC -3') containing an internal PstI site was designed to anneal upstream to the EcoRV cloning site of pT7Blue. A reverse primer (5'-CgCCTgCAggTAATACgACTCACTATAggg -3') also containing a PstI site was designed based on pBluecsript phagemid sequence immediately 3' to the EcoRI/XhoI cloning site. Full length PERK1 cDNA was generated from a 100 microliter PCR reaction containing 1 microliter (~20ng) of each template (cDNA in pT7Blue and pBluescript phagemid), 10x Pfu Buffer (200mM Tris-HCl pH8.8, 100mM $(NH_4)_2SO_4$, 20mM $MgSO_4$, 1% Triton®X-100, 1mg/mlBSA), 10mM dNTPs, 50pmol forward and reverse primers and 1microliter Pfu polymerase (Gibco-BRL, Gainthersburg, MD). The samples were heated to 94°C for 5 min and amplified for 30 cycles with a denaturing cycle of 1 min, a pimer annealing cycle at 53°C for 1 min followed by an extension cycle for 3 min at 72°C. The resulting PCR product of the expected size (~2.2kb) was gel purified and cloned into the PstI restriction site of pBluescript KS (+/-) II. The full length PERK1 cDNA sequence was confirmed by a sequential primer based sequencing approach using both universal and sequence specific primers as previously described. All DNA and protein sequence analysis was performed using the DNAsis® Software (Hitachi Software, San Bruno,CA).

Genomic DNA Isolation and Southern Blot Analysis

Genomic DNA was extracted from approximately one gram of young *Brassica napus* leaf tissue according to the method described by Goring et al. (1992b). Approximately 5:g of genomic DNA was digested with several restriction enzymes (BamHI, EcoRI, HindIII, PstI, XbaI, XhoI), fractionated through a 0.8% agarose gel and transferred overnight in 10x SSC onto Zetaprobe membrane (Biorad, Hercules,CA). This was performed in duplicate to test hybridization conditions under low and high stringencies conditions. After drying, the

membranes were prewashed in 0.1x SSC, 0.5% SDS for 25 min at 60°C. The membranes were prehybridized and hybridized as previously described for plasmid Southern blots with the inclusion of 10% dextran sulfate and 50 microgram/milliliter salmon sperm DNA in the hybridization buffer. Washing conditions for genomic southern blots varied depending on the stringency tested. One membrane was washed under conditions of low stringency for 15 min at room temperature in 2x SSC, 0.1% SDS followed by second 15 min room temperature wash in 1x SSC, 0.1% SDS and three final washes at 50°C in 1x SSC, 0.1% SDS.. The second membrane was washed under conditions of high stringency by lowering the salt concentration to 0.1x SSC, 0.1% SDS and increasing the temperature to 65°C. The ^{32}P-labeled 1512 bp PERK1 cDNA probe was generated by random priming as described by Feinberg and Vogelstein (1983). Membranes were subjected to autoradiography (XAR-5 film, Kodak) overnight at -80°C.

Isolation and Northern Blot Analysis of Multiple Tissue RNA

Total RNA was extracted from a mixture of Westar and W1 root, stem, leaf, petal, anther and pistil tissue as described by Jones et al. (1985). Poly(A)$^+$ mRNA was isolated using the polyA Spin™ mRNA Isolation kit as outlined by the manufacturer's procedure (New England Biolabs, Beverly, MA). Approximately 3 micrograms of poly(A)$^+$ mRNA was fractionated on a 1.2% formaldehyde gel (Sambrook et al., 1989) and transferred to Zetaprobe membrane (Biorad, Hercules, CA) in 10x SSC. Hybridization and high stringency wash conditions were conducted as previously described for genomic Southern blot analysis. The membrane was subsequently probed with a cyclophilin EST clone (No. mBN086) as an internal control for even loading of poly(A)$^+$ mRNA.

Plant Treatments

B. napus plants were grown in a growth chamber at 22°C with a 16hr light period followed by an 8hr dark period at 16°C. Experiments were conducted on two month old plants, and all experiments used one plant per time point from which leaf and stem tissue was harvested.

Wounding of leaf material was performed by punching out leaf discs every 1cm around the perimeter of the leaf blade ensuring that the midvein remained intact, and stems were wounded by slicing into 1-3cm segments. The wounded tissues were placed in petri dishes

containing filter paper moistened with 20mM sodium phosphate buffer supplemented with 50 microgram/milliliter chloramphenicol to prevent bacterial contamination of the wounded tissue (Shirsat et al., 1996). A control (0 hr) time point for this experiment was performed by incubating unwounded tissue in the sodium phosphate buffer. Wounded leaf and stem tissue was harvested at different times after wouding (0 hr, 5 min, 15 min, 45 min, 1 hr, 4 hr, 12 hr, 24 hr, 36 hr, 48 hr).

Plants were thoroughly sprayed with 50 micromolar methyl jasmonate (MeJA; Sigma, St. Louis, MO) (Titarenko et al., 1997) and 4mM salicylic acid (SA; Sigma, St. Louis, MO) solutions (Schweizer et al., 1998). Leaf and stem tissue was harvested at various time points (0 hr, 5 min, 15 min, 30 min, 45 min, 1 hr, 2 hr, 4 hr, 12 hr, 24 hr, 36 hr, 48 hr, 72 hr, 96 hr) following SA and MeJA treatments. Control sprays were performed with the carrying solutions in the absence of the chemical inducer. Carrying solutions were 5mM phosphate buffer, pH 7 for SA, and 0.1% [%$_v$] ethanol for MeJA.

Total RNA was extracted from treated tissue according to the method described by Cock et al. (1997). Depending on the treatment, varying amounts of total RNA (20-40 micrograms) was electrophoresed on a 1.2% formaldehyde and standard Northern blot analysis was performed as described by Sambrook et al. (1989). Hybridization and washing conditions were performed as outlined for the multiple tissue northern blot. Following autoradiography, the amounts of radioactive signal were quantified using Instant Imager Electronic Autoradiography (Packard, Meriden, CT). The membranes were reprobed with the cyclohpilin cDNA and the amounts of hybridized radiolabeled cyclophilin were quantified in the same manner. The relative amounts of RNA hybridized to the full length PERK1 cDNA probe were determined after correction for differences in the amounts of cyclophilin RNA.

The present invention has been described in detail and with particular reference to the preferred embodiments; however, it will be understood by one having ordinary skill in the art that changes can be made thereto without departing from the spirit and scope of the invention.

All articles, patents and other documents described in this application (including database sequences and/or accession numbers), US patent application no. 60/149,466 filed on August 19,1999 (entitled "Characterization of a Novel Receptor Kinase from *Brassica* with a Putative

Role in Plant Defence") and US patent nos. 5,612,191, 5,763,211, 5,750,848 and 5681714, are incorporated by reference in their entirety to the same extent as if each individual publication, patent or document was specifically and individually indicated to be incorporated by reference in its entirety. They are also incorporated to the extent that they supplement, explain, provide a background for, or teach methodology, techniques and/or compositions employed herein.

REFERENCES

Braun, D.M., and Walker, J.C. (1996). Plant transmembrane receptors: new pieces in the signalling puzzle. TIBS **21,** 70-73.

Cassab, G. I. (1998). Plant cell wall proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49,** 281-309.

Chang, C., Schaller, G.E., Patterson, S.E., Kwok, S.F., Meyeroxitz, E.M., and Bleeker, A.B. (1992). The TMK1 gene from Arabidopsis codes for a protein with structural and biochemical characteristics of a receptor protein kinase. Plant Cell **4,** 1263-1271.

Chen, J., and Varner, J.E. (1985). An extracellular matrix protein in plants: Characterization of a genomic clone for carrot extensin. EMBO J. **4,** 2145-2151.

Clark, S., Williams, R.W., and Meyerowitz, E.M. (1997). The *CLAVATA1* gene encodes a putative receptor-kinase that controls shoot and floral meristem size in *Arabidopsis*. Cell **89,** 575-585.

Cock, J.M., Swarup, R., and Dumas, C. (1997). Natural antisense transcripts of the S locus receptor kinase gene and related sequences in Brasscia oleracea. Mol. Gen. Gent. **255,** 514-524.

Corbin, D.R., Sauer, N., Lamb, C.J. (1987). Differential regulation of a hydroxyproline-rich glycoprotein gene family in wounded and infected plants. Mol. Cell Biol. **7,** 4337-4344.

Doares, S.H., Narváez-Vásquez, J., Conconi, A., and Ryan, C.A. (1995). Salicylic acid inhibits synthesis of proteinase inhibitors in tomato leaves inducted by systemin and jasmonic acid. Plant Physiol. **108**, 1741-1746.

Dwyer, K.G., Kandasamy, M.K., Mahosky, D.I. Acciai, J., Kudish, B.I. Miller, J.E., Nasrallah, M.E., and Nasrallah, J.B. (1994). A superfamily of S locus-related sequences in *Arabidopsis*: Diverse structures and expression patterns. Plant Cell **6**, 1829-1843.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**, 6-13.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full length DNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85**, 8998-9002.

Gasser, C.S., Gunning, D.A., Budelier, K.A., and Brown, S.M. (1990). Structure and expression of cytosolic cyclophilin/peptidyl-proylcis-trans isomerase of higher plants and the production of active tomato cyclophilin in *Eschericia coli*. Proc. Natl. Acad. Sci. USA **87**, 9519-9523.

Goring, D.R., and Rothstein, S.J. (1992). The S-locus receptor kinase gene in a self-incompatible *Brassica napus* line encodes a functional serine/threonine kinase. Plant Cell **4**, 1273-1281.

Goring, D.R., Banks, P., Fallis, L., Baszczynski, C.L., Beversdorf, W.D., and Rothstein, S.J. (1992b). Identification of an S-locus glycoprotein allele introgressed form B. napus ssp. oleifera. Plant J. **2**, 983-989.

Hanks, S.K., and Quinn, A.M. (1991). Protein kinase catalytic domain sequence database: Identification of conserved features of 1° structure and classification of family members. Methods Enzymol. **200,** 38-62.

Hervé, C., Dabos, P. Galaud, J.-P. Rougé, P., and Lescure, B. (1996). Characterization of an *Arabidopsis thaliana* gene that defines a new class of putative plant receptor kinases with an extracellular lectin-like domain. J. Mol. Biol. **258,** 778-788.

Horn, M.A., and Walker, J.C. (1994). Biochemical properties of the autophosphorylation of RLK5, a receptor-like protein kinase from *Arabidopsis thaliana.* Biochem. et Biophys. Acta **1208,** 65-74.

Jones, J.D.G., Dunsmuir, P., and Bedbrook, J. (1985). High level expression of introduced chimeric genes in regenerated transformed plants. EMBO **4,** 2411-2418.

Keller, B., and Lamb, C.J. (1989). Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. Genes Dev. **3,** 1639-1646.

Kohorn, B.D., Lane, S., and Smith, T.A. (1992). An *Arabidopsis* serine/threonine kinase homologue with an epidermal growth factor repeat selected in yeast for its specificity for a thylakoid membrane protein. Proc. Natl. Acad. Sci. USA **89,** 10989-10992.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157,** 105-132.

Li, J., and Chroy, J. (1997). A putative leucine rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell **90,** 929-938.

Lin, H.V., Wang, X.F., Ng-Eaton, E., Weinberg, R.A., and Lodish, H.F. (1992). Expression cloning of TGF-type II receptor, a functional transmembrane serine/threonine kinase. Cell **68,** 775-785.

Lutcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., and Scheele, G.A. (1987). Selection of AUG initiation codons differs in plants and animals. EMBO J. **6,** 43-48.

Merkouropoulos, G., Barnett, D.C., and Shirsat, A.H. (1999). The *Arabidopsis* extensin gene is developmentally regulated, is induced by wounding, methyl jasmonate, abscissa acid and salicylic acid, and coded for a protein with unusual motifs. Planta **208,** 212-219.

Mu, J.-H., Lee, H.-S., and Kao, T. (1194). Characterization of a pollen expressed receptor-like kinase gene of *Petunia inflata* and the activity of its encoded kinase. Plant Cell **6,** 709-721.

Nasrallah, J.B., and Nasrallah, M.E. (1993). Pollen-stigma signaling in the sporophytic self-incompatibility response. Plant Cell **5,** 1325-1335.

Penninckx, I.A.M., Eggermont, K., Terras, F.R.G., Thomma, B.P. Samblanx, G.W. Buchala, A., Métraux, J.P., Manners, J.M. and Broekaert, W.F. (1996). Pathogen-induced systemic activation of a plant defensin gene in *Arabidopsis* follows a salicylic acid-independent pathway. Plant Cell **8,** 2309-2323.

Peña-Cortés, H., Albercht, T., Prat, S., Weiler, E.W., and Willmitzer, L. (1993). Aspirin prevents wound-induced gene expression in tomato leaves by jasmonic acid biosynthesis. Planta **191,** 123-128.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: A Laboratory Manual, 2nd ed. New York: Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74,** 5463-5467.

Sauer, N., Corbin, D.R., Keller, B., and Lamb, C.J. (1990). Cloning and characterization of a wound specific hydroxyproline-rich glycoprotein in *Phaseolus vulgaris*. Plant Cell Environ. **13,** 257-266.

Schweizer, P., Buchala, A., Dudker, R., and M9traux, J.P. (1998). Induced systemic resistance in wounded rice plants. Plant J. **14,** 475-481.

Shirsat, A.H., Wieczoreky, D., and Kozibal, P. (1996). A gene for Brassica napus extensin is differentially expressed on wounding. Plant Mol. Biol. **30,** 1291-1300.

Showalter, A.M., Zhou, J., Rumeau, D., Worst, S.G., and Varner, J.E. (1991). Tomato extensin and extensin-like cDNAs: Stucture and expression in response to wounding. Plant Mol. Biol. **16,** 547-565.

Showlater, A.M. (1993) Structure and function of plant cell wall proteins. Plant Cell **5,** 9-23.

Singer, S. J. (1990). The structure and insertion of integral proteins in the membrane. Annu. Rev. Cell Biol. **6,** 247-296.

Song, W.-Y., Wang, G.-L., Chen, L., Kim, H.-S., Pi, L.-Y., Gardner, J., Wang, B., Holsten, T., Zhai, W.-X.., Zhu, L.-H., Fauquet, C., and Ronald, P.C. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene *Xa21*. Science **270,** 1804-1806.

Stein, J.C., Howlett, B., Boyes, D.C., Nasrallah, M.E., and Nasrallah, J.B. (1991). Molecular cloning of a puatitve receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. Proc. Natl. Acad. Sci. USA **88,** 8816-8820.

Suzuki, K., and Shinshi, H. (1995). Transient activation and tyrosine phosphorylation of a protein kinase in tobacco cells treated with a fungal elicitor. Plant Cell **7,** 639-647.

Swarup, R., Dumas, D., and Cock, J.M. (1996). A new class of receptor-like protein kinase gene from *Arabidopsis thalaina* possessing a domain with similarity of plant lectin genes. Plant Physiol. **111,** 347.

ten Dijke, P., Ichijo, H., Franzen, P., Schulz, P., Saras, J., Toyoshima, H., Heldin, C., and Miyazono, K. (1993). Activin receptor-like kinases: A novel subclass of cell surface receptors with predicted serine/threonine kinase activity. Oncogene **8,** 2879-2887.

Titarenko, E., Rojo, E., Léon, J., and Sanchez-Serrano, J. (1997). Jasmonic acid - dependent and -independent signalling pathways control wound-induced gene activation in Arabidopsis thaliana. Plant Physiol. **115,** 817-826.

Tobias, C.M., Howlett, B., and Nasrallah, J.B. (1992). An *Arabidopsis thaliana* gene with sequence similarity to the S-locus receptor kinase of *Brassica oleracea*. Plant Physiol. **99,** 284-290.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F. and Komeda, Y. (1996). The Arabidopsis *ERECTA* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant Cell **8,** 735-746.

Truernit, E., Schmid, J., Epple, P., Illig, J., and Sauer, N. (1996). The sink specific stress-regulated Arabidopsis *STP4* gene: Enhanced expression of a gene encoding a monosaccharide transporter by wounding, elicitors and pathogen challenge. Plant Cell **8,** 2169-2182.

Ullrich, A., and Schlessinger, J. (1990). Signal transduction by receptors with tyrosine kinase activity. Cell **61,** 203-212.

Usami, S., Banno, H. Ito, Y., Nishihama, R., and Machida, Y. (1995). Cutting activates a 46-kilodalton protein kinase in plants. Proc. Natl. Acad. Sci. USA **92,** 8660-8664.

Walker, J.C. (1993). Receptor-like protein kinase genes of *Arabidopsis thaliana*. Plant J. **3,** 451-456.

Walker, J.C. (1994). Structure and function of the receptor-like protein kinases of higher plants. Plant Mol. Biol. **26,** 1599-1609.

Walker, J.C., and Zhang, R. (1990). Relationship of a putative receptor protein kinase form maize tot he S-locus glycoproteins of *Brassica*. Nature **345,** 743-746.

Wang, G.-L., Ruan, D.-L., Song, W.-Y., Sideris, S., Chen, L., Pi, L.-Y., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B.S., Whaten, M.C., and Ronald, P.C. (1998). Xa21 encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell **10,** 765-779.

Wang, X., Zafian, P., Choudhary, M., and Lawton, M. (1996). The PR5K receptor protein kinase form *Arabidopsis thaliana* is structurally related to a family of plant defense proteins. Proc. Natl. Acad. Sci. USA **93,** 2598-2602.

Ward, .R., Uknes, S.J., Williams, S.C., Dincher, S.S., Wiederhold, D.L., Alexander, D.C., Ahl-Goy, P., Métraux, J.-P., and Ryals, J.A. (1991). Coordinate gene activity in response to agents that induce systemic acquired resistance. Plant Cell **3,** 1085-1094.

Weinstein, J.N., Blumenthal, R., Van Renwoude, J., Kempf, C., and Mauser, R.O. (1982). Charge clusters and the orientation of membrane proteins. J. Membr. Biol. **66,** 203-212.

Wilson, L.G., and Fry, J.L. (1986). Extensins: A major cell wall glycoprotein. Plant Cell Environ. **9,** 239-260.

Yang, Y., Shah, J., and Klessig, D.F. (1997). Signal perception and transduction in plant defense responses. Genes Dev. **11,** 1621-1639.

Zhao, Y., Feng, X.-H., Watson, J.C., Bottino, P.J., and Kung, S.-D. (1994). Molecular cloning and biochemical characterization of a receptor-like serine/threonine kinase from rice. Plant Mol. Biol. **26,** 791-803.

Zhou, J., Loh, Y.-T., Bressan, R.A., and Martin, G.B. (1995). The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response. Cell **83,** 925-935.

We claim:

1. An isolated nucleic acid molecule encoding a proline-rich, extensin-like receptor kinase (PERK) polypeptide, or a fragment of a PERK polypeptide having PERK activity.

2. The molecule of claim 1, wherein the polypeptide is a signaling molecule associated with the cell wall via its extensin-like extracellular domain and is involved in the transduction of extracellular stimuli into an intracellular response through a cytoplasmic kinase domain, thereby bridging the cell wall-plasma membrane continuum.

3. The molecule of claim 1, wherein the extracellular stimuli includes wounding or pathogen attack.

4. The molecule of claim 1, wherein the wounding is selected from the group consisting of a cut, a break, a tear, a fold and an insect wound.

5. The molecule of claim 3, wherein the pathogen comprises bacterial pathogens, fungal pathogens, *Sclerotinia sclerotiorum, Cylindrosporium concertricum, Phoma lingam, Pseudomonas syringae, Streptomyces scabies,* Blackleg, Whiterust, *Fusarium* Head Blight, Rust, Bunt, Leaf Spot , White mold, root rot, *Fusarium* ear rot

6. An isolated nucleic acid molecule encoding a PERK polypeptide, a fragment of a PERK polypeptide having PERK activity, or a polypeptide having PERK activity, comprising a nucleic acid molecule selected from the group consisting of:

   (a) a nucleic acid molecule that hybridizes to all or part of a nucleic acid molecule shown in [SEQ ID NO:1], or a complement thereof under low, moderate or high stringency hybridization conditions wherein the nucleic acid molecule encodes a PERK polypeptide or a polypeptide having PERK activity;

   (b) a nucleic acid molecule degenerate with respect to (a), wherein the nucleic molecule encodes a PERK polypeptide or a polypeptide having PERK activity.

7. The nucleic acid molecule of claim 2, wherein the hybridization conditions comprise low stringency conditions of 1XSSC, 0.1% SDS at 50°C or high stringency conditions of 0.1XSSC, 0.1% SDS at 65°C.

8. An isolated nucleic acid molecule encoding a PERK polypeptide, a fragment of a PERK polypeptide having PERK activity, or a polypeptide having PERK activity, comprising a nucleic acid molecule selected from the group consisting of:

   (a) the nucleic acid molecule of the coding strand shown in [SEQ ID NO:1], or a complement thereof;

   (b) a nucleic acid molecule encoding the same amino acid sequence as a nucleotide sequence of (a); and

   (c) a nucleic acid molecule having at least 17% identity with the nucleotide sequence of (a) and which encodes a PERK polypeptide or a poypeptide having PERK activity

9. The nucleic acid molecule of any of claims 1 to 8, wherein the PERK polypeptide comprises a PERK1 polypeptide.

10. The nucleic acid molecule of claim 1, comprising all or part of a nucleotide sequence shown in [SEQ ID NO:1] or a complement thereof.

11. A PERK1 nucleic acid molecule isolated from *Brassica*, or a fragment thereof.

12. The nucleic acid molecule of claim 11, wherein the *Brassica* comprises *Brassica napu,s Brassica juncea, Brassica rapa or Brassica oleracea.*

13. A recombinant nucleic acid molecule comprising a nucleic acid molecule of any of claims 1 to 4 and a constitutive promoter sequence or an inducible promoter sequence, operatively linked so that the promoter enhances transcription of the nucleic acid molecule in a host cell.

14. The nucleic acid molecule of claim 1, wherein the molecule comprises genomic DNA, cDNA or RNA .

15. The nucleic acid molecule of claim 14, wherein the nucleic acid molecule is chemically synthesized.

16. An isolated nucleic acid molecule comprising a nucleic acid molecule selected from the group consisting of 8 to 10 nucleotides of the nucleic acid molecule of claim 6, 11 to 25 nucleotides of the nucleic acid molecule of claim 6 and 26 to 50 nucleotides of the nucleic acid molecule of claim 10.

17. A vector comprising the nucleic acid molecule of any of claims 1 to 4.

18. The vector of claim 17, comprising a promoter selected from the group consisting of a super promoter, a 35S promoter of cauliflower mosaic virus, a chemical inducible promoter, a a copper-inducible promoter, a steroid-inducible promoter and a tissue-specific promoter.

19. A host cell comprising the recombinant nucleic acid molecule of claim 1 or the vector of claim 17, or progeny of the host cell.

20. The host cell of claim 19, selected from the group consisting of a fungal cell, a yeast cell, a bacterial cell, a microorganism cell and a plant cell.

21. A plant, a plant part, a seed, a plant cell or progeny thereof comprising the recombinant nucleic acid molecule of claim 13 or the vector of claim 17.

22. The plant part of claim 21, comprising all or part of a leaf, a flower, a stem, a root or a tuber.

23. The plant, plant part, seed or plant cell of claim 21, wherein the plant, plant part, seed or plant cell is of a species selected from the group consisting of *Brassica napus, Brassica rapa, Brassica juncea. Brassica oleracea*, or from the family Brassicaecae, Arabidopsis, potato, tomato, tobacco, cotton, carrot, petunia, sunflower, strawberries, spinach, lettuce, rice, soybean, corn, wheat, rye, barley, sorgum and alfalfa.

24. The plant, plant part, seed or plant cell of claim 21, wherein the plant comprises a dicot plant.

25. The plant, plant part, seed or plant cell of claim 21, wherein the plant comprises a monocot plant.

26. An isolated polypeptide encoded by and/or produced from the nucleic acid molecule of any of claims 1 to 4, or the vector of claim 17.

27. An isolated PERK polypeptide or a fragment thereof having PERK activity.

28. The polypeptide of claim 27 comprising all or part of an amino acid sequence in [SEQ ID NO:1].

29. A polypeptide fragment of the PERK polypeptide of claim 27, or a peptide mimetic of the PERK polypeptide.

30. The polypeptide fragment of claim 28, consisting of at least 20 amino acids, which fragment has PERK activity.

31. The fragment or peptide mimetic of claim 29, which is capable of being bound by an antibody to the polypeptide of claim 28.

32. The polypeptide of claim 26 which is recombinantly produced.

33. An isolated and purified polypeptide comprising the amino acid sequence of a PERK polypeptide, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes under moderate or stringent conditions to a nucleic acid molecule in [SEQ ID NO:1], a degenerate form thereof or a complement.

34. A polypeptide comprising a sequence having greater than 20% sequence identity to the polypeptide of claim 28.

35. The polypeptide of claim 26, wherein the polypeptide comprises a PERK polypeptide.

36. The polypeptide of claim 35, isolated from *Brassica*.

37. The polypeptide of claim 36, wherein the *Brassica* comprises *Brassica napus* or *Brassica juncea* or *Brassica rapa or Brassica oleracea.*

38. The polypeptide of claim 34, comprising a kinase domain including at least 30% homology to the kinase domain of [SEQ ID NO.:2] and/or an extracellular domain including at least 20% homology to the extracellular domain of [SEQ ID NO.:2].

39. An isolated nucleic acid molecule encoding the polypeptide of claims 27 or 28.

40. An antibody directed against the polypeptide of claim 28.

41. The antibody of claim 40, comprising a monoclonal antibody or a polyclonal antibody.

42. An assay to identify a polypeptide having PERK activity, comprising:

43. An isolated nucleic acid molecule encoding a polypeptide that reduces the severity of wounding or pathogen attack in a plant, the polypeptide comprising:

(a) an extracellular domain which recognizes an extracellular binding molecule whose level is increased during the wounding or pathogen attack, the extracellular domain encoding a plurality of repeats selected from the group consisting of SPPPP, SPP, PP and PPP, wherein a plurality of the proline molecules are capable being glycosylated and/or hydroxylated;

(b) a membrane domain operably connected to the extracellular domain, wherein the membrane domain is capable of extending across a cell membrane from the extracellular side of the membrane to intracellular side of the membrane; and

(c) a cytoplasmic domain operably connected to the membrane domain, wherein the cytoplasmic domain comprises a means for producing kinase activity when the extracellular binding molecule interacts with the extracellular domain.

ABSTRACT

The invention includes PERK (Proline-rich Extensin-like Receptor Kinase) nucleic acid molecules and polypeptides. A receptor-like protein kinase designated PERK1 (Proline-rich Extensin-like Receptor Kinase 1) was isolated from a 8-pistil cDNA library of *Brassica napus*. The deduced PERK1 protein is comprised of a cytoplasmic domain that contains all of the conserved amino acids prevalent among serine/threonine kinases, a transmembrane domain and an extracellular domain with sequence similarity to the extensin family of plant cell wall proteins. Northern blot analysis demonstrated that PERK1 mRNA accumulated rapidly in leaf and stem tissue of *B. napus* in response to wounding and treatment with salicylic acid. In contrast, no significant accumulation of PERK1 mRNA was detected following treatment with methyl jasmonate. The rapidity of PERK1 mRNA accumulation in response to these treatments shows a role in plant defense signaling.

PTO/SB/09 (12-97)
Approved for use through 09/30/00. OMB 0651-0031
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT CLAIMING SMALL ENTITY STATUS (37 CFR 1.9 (f) & 1.27 (b)) –INDEPENDENT INVENTOR	Docket Number (Optional) 2846/0002

Applicant, Patentee, or Identifier: SILVA, Nancy F. and GORING, Daphne R.

Application or Patent No.: _____

Filed or Issued: _____

Title: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES

As a below named inventor, I hereby state that I qualify as an independent inventor as defined in 37 CFR 1.9(c) for purposes of paying reduced fees to the Patent and Trademark Office described in:

☒ the specification filed herewith with title as listed above

☐ the application identified above.

☐ the patent identified above.

I have not assigned, granted, conveyed, or licensed, and am under no obligation under contract or law to assign, grant, convey, or license, any rights in the invention to any person who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person had made the invention, or to any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or an nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization to which I have assigned, granted, conveyed, or licensed or am under an obligation under contract or law to assign, grant, convey, or license any rights in the invention is listed below:

☒ No such person, concern, or organization exists.

☐ Each such person, concern, or organization is listed below.

*NOTE: Separate statements are required from each named person, concern, or organization having rights to the invention stating their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patents to which this verified statement is directed.

SILVA, Nancy F.	GORING, Daphne R.	
NAME OF INVENTOR	NAME OF INVENTOR	NAME OF INVENTOR
Nancy Fonseca Silva	Daphne Goring	
Signature of inventor	Signature of inventor	Signature of inventor
Oct. 12, 1999.	Oct. 12, 1999	
Date	Date	Date

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual cases. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

A.

```
                                                  CCTCGGCGCCGTCTCCGGGG    24
                                                    S  S  A  P  S  P  G
  25 ACTGGTTCGCCTCCATCTCCACCATCAAACTCCACAACCACCACTCCTCCTCCAGCTTCC      84
     T  G  S  P  P  S  P  P  S  N*  S  T  T  T  T  P  P  P  A  S
  85 GCTCCTCCTCCCACCACACCTTCTTCTCCTCCGCCGCCATCCACTATTCCGACATCTCCT     144
     A  P  P  P  T  T  P  S  S  P  P  P  P  S  T  I  P  T  S  P
 145 CCTCCTTCTTCTCGCTCTACACCTTCTGCTCCTCCTCCATCTCCACCAACTCCATCTACG     204
     P  P  S  S  R  S  T  P  S  A  P  P  P  S  P  P  T  P  S  T
 205 CCGGGATCTCCACCTCCTCTTCCTCAGCCGTCTCCACCCGCTCCAACTACGCCCGGATCT     264
     P  G  S  P  P  P  L  P  Q  P  S  P  P  A  P  T  T  P  G  S
 265 CCACCCGCACCTGTTACTCCTCCTACTCGAAACCCTCCACCTTCAGTCCCAGGACCACCG     324
     P  P  A  P  V  T  P  P  T  R  N  P  P  P  S  V  P  G  P  P
 325 TCCAATCCTTCACGCGAAGGAGGATCTCCTCGACCTCCATCTTCTCCCTCGCCGCCGTCT     384
     S  N  P  S  R  E  G  G  S  P  R  P  P  S  S  P  S  P  P  S
 385 CCTTCTTCCGACGGTTTATCAACAGGAGTGGTGGTGGGAATCGCCATCGGAGGAGTCGCT     444
     P  S  S  D  G  L  S  T  G  V  V  V  G  I  A  I  G  G  V  A
 445 CTGCTTGTGATAGTGACTCTGATTTGTCTCCTCTGTAAGAAGAAACGACGGAGAGACGAA     504
     L  L  V  I  V  T  L  I  C  L  L  C  K  K  K  R  R  R  D  E
 505 GAAGATGCTTACTATGTTCCTCCGCCACCTCCTCCTGGTCCCAAAGCCGGAGGACCTTAC     564
     E  D  A  Y  Y  V  P  P  P  P  P  P  G  P  K  A  G  G  P  Y
 565 GGTGGACAGCAGCAACAATGGCGGCAACAAAACGCAACACCACCGTCAGATCATGTCGTG     624
     G  G  Q  Q  Q  Q  Q  W  R  Q  Q  N  A  T  P  P  S  D  H  V  V
 625 ACGTCACTACCACCACCACCTAAGGCTCCATCTCCACCACGGCAACCTCCTCCACCTCCA     684
     T  S  L  P  P  P  P  K  A  P  S  P  P  R  Q  P  P  P  P  P
 685 CCACCGCCTTTCATGAGCAGCAGCGGCGGCTCCGACTACTCGGACCGTCCAGTTCTTCCT     744
     P  P  P  F  M  S  S  S  G  G  S  D  Y  S  D  R  P  V  L  P
 745 CCACCGTCTCCAGGGCTTGTGTTAGGCTTCTCCAAAAGCACTTTCACATACGAGGAGCTA     804
     P  P  S  P  G  L  V  L  G  F  S  K  S  T  F  T  Y  E  E  L
 805 GCTAGAGCCACCAATGGTTTCTCCGAGGCGAACTTGTTAGGACAAGGCGGGTTCGGTTAC     864
     A  R  A  T  N  [3]G  F  S  E  A  N  L  L  G  Q  G  G  F  G  Y
 865 GTGCACAAAGGTGTGTTGCCTAGTGGGAAAGAAGTTGCTGTGAAGCAGTTGAAAGTTGGG     924
     V  H  K  G  V  L  P  S  [11]G  K  E  V  A  V  K  Q  L  K  V  G
 925 AGTGGTCAGGGAGAGAGGGAGTTTCAGGCAGAGGTTGAGATCATCAGCAGAGTTCACCAC     984
     S  G  [111]Q  G  E  R  E  F  Q  A  E  V  E  I  I  S  R  V  [IV]H  H
 985 AGGCATCTGGTGTCTCTTGTTGGTTATTGCATCGCCGGTGCCAAAAGATTGCTTGTCTAT    1044
     R  H  L  V  S  L  V  G  Y  C  I  A  [V]G  A  K  R  L  L  V  Y
1045 GAGTTTGTTCCTAACAACAATCTCGAGCTTCACCTCCATGGCGAGGGACGGCCTACAATG    1104
     E  F  V  P  N  N  N  L  E  L  H  L  H  G  E  G  R  [VI]P  T  M
1105 GAATGGAGCACCAGATTGAAGATTGCTCTTGGATCTGCTAAAGGACTTTCTTATCTTCAT    1164
     E  W  S  T  R  L  K  I  A  L  G  S  A  K  G  L  S  Y  L  H
1165 GAAGATTGCAATCCTAAAATCATTCACCGTGATATCAAGGCTTCAAACATATTGATAGAT    1224
     E  D  C  N  P  K  I  I  H  R  D  I  K  A  S  N  I  L  I  D
1225 TTCAAGTTTGAAGCTAAGGTTGCTGATTTTGGTCTTGCTAAGATTGCTTCTGATACAAAC    1284
     F  K  F  [VII]E  A  K  V  A  D  F  G  L  A  K  I  A  S  D  T  N
1285 ACGCATGTATCAACACGTGTGATGGGAACCTTTGGGTACTTGGCTCCGGAATACGCTGCA    1344
     T  H  V  [VIII]S  T  R  V  M  G  T  F  G  Y  L  A  P  E  Y  [IX]A  A
1345 AGCGGAAAGCTCACGGAGAAGTCTGACGTTTTCTCATTTGGCGTTGTGCTTTTGGAGCTC    1404
     S  G  K  L  T  E  K  S  D  V  F  S  F  G  V  V  L  L  F  L
1405 ATTACTGGACGTCGACCCGTTGATGCCAACAATGTCTATGTAGATGACAGCTTAGTTGAC    1464
     I  T  G  R  R  P  V  D  A  N  N  [X]V  Y  V  D  D  S  L  V  D
1465 TGGGCACGACCATTGCTTAACCGAGCATCTGAGCAAGGAGACTTTGAGGGTTTAGCTGAT    1524
     W  A  R  P  I  L  N  R  A  S  E  Q  G  D  E  F  G  L  A  D
1525 GCAAAGATGAATAATGGGTATGACAGAGAGGAGATGGCTCGCATGGTTGCTTGTGCTGCG    1584
     A  K  M  N  N  G  Y  D  R  E  [XI]E  M  A  R  M  V  A  C  A  A
1585 GCTTGTGTTCGCCATTCAGCTCGCCGCAGACCTCGCATGAGCCAGATTGTGCGTGCGTTA    1644
     A  C  V  R  H  S  A  R  E  R  P  R  M  S  Q  I  V  R  A  L
1645 GAAGGAAATGTATCACTGTCAGATCTTAACGAAGGGATGAGACCAGGTCAAAGCAATGTA    1704
     E  G  N*  V  S  L  S  D  L  N  E  G  M  R  P  G  Q  S  N  V
1705 TACAGCTCATACGGAGGAAGCACCGATTATGACTCGAGCCAGTACAATGAAGACATGAAG    1764
     Y  S  S  Y  G  G  S  T  D  Y  D  S  S  Q  Y  N  E  D  M  K
1765 AAGTTTAGGAAAATGGCACTTGGAACTCAAGAGTACAACGCCACGGGTGAGTACAGTAAT    1824
     K  F  R  K  M  A  L  G  T  Q  E  Y  N*  A  T  G  E  Y  S  N
1825 CCGACCAGTGACTATGGACTGTACCCGTCTGGTTCAAGCAGCGAGGGCCAAACCACACGC    1884
     P  T  S  D  Y  G  L  Y  P  S  G  S  S  S  E  G  Q  T  T  R
1885 GAAATGGAGATGGGGAAGATTAAGAGAACCGGTCAGGGTTATAGTGGACCTTCTCTTTAA    1944
     E  M  E  M  G  K  I  K  R  T  G  Q  G  Y  S  G  P  S  L  *
```

B.

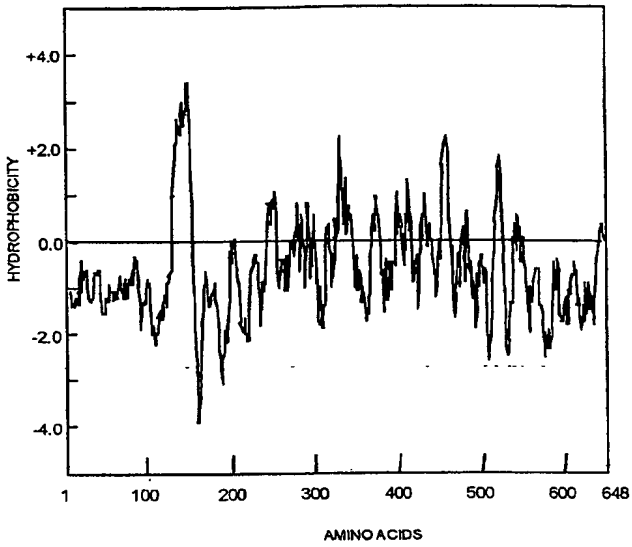


Figure 1

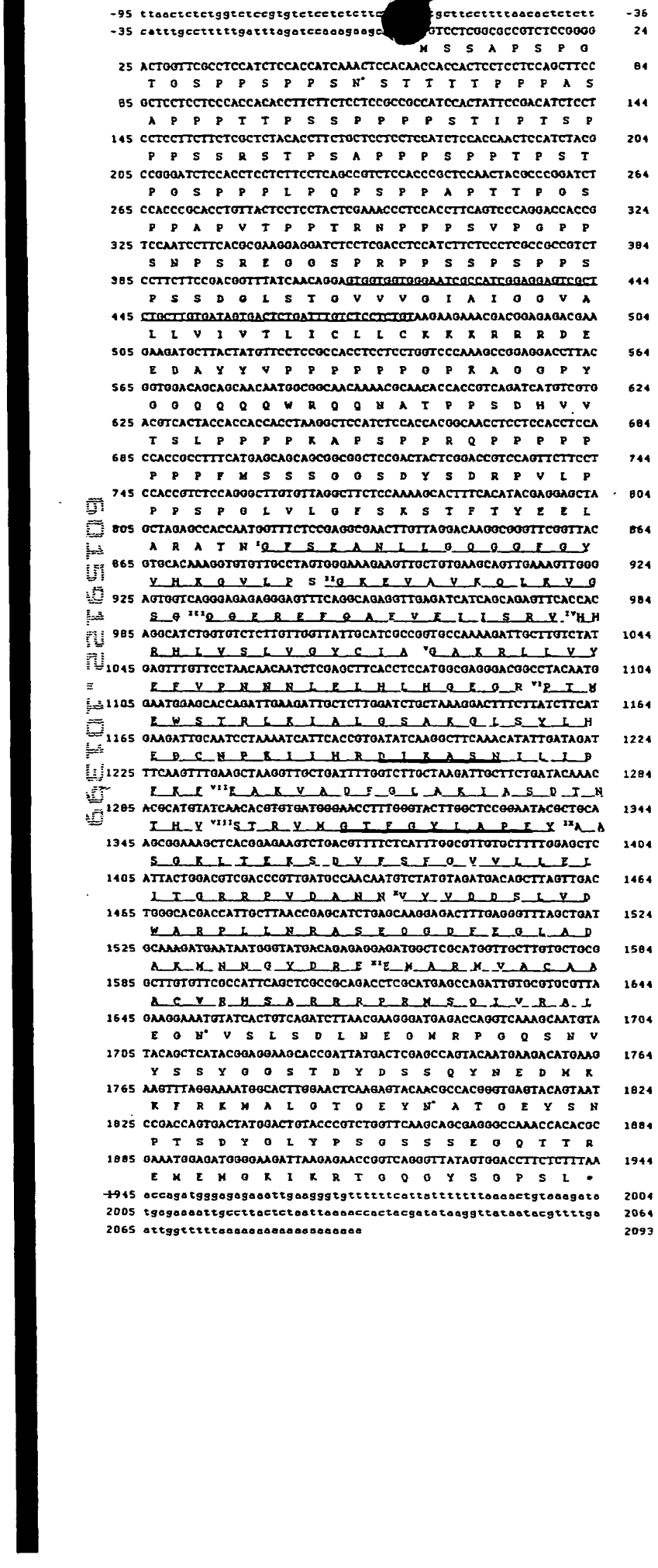


Figure 1C

```
        10          20          30          40          50          60
TTAACTCTCT GGTCTCCGTG TCTCCTCTCT TCTCCTGCTG CTTCCTTTTA ACACTCTCTT
        70          80          90         100         110         120
CATTTGCCTT TTTGATTTAG ATCCAAAGAA GCAGACATGT CCTCGGCGCC GTCTCCGGGG
       130         140         150         160         170         180
ACTGGTTCGC CTCCATCTCC ACCATCAAAC TCCACAACCA CCACTCCTCC TCCAGCTTCC
       190         200         210         220         230         240
GCTCCTCCTC CCACCACACC TTCTTCTCCT CCGCCGCCAT CCACTATTCC GACATCTCCT
       250         260         270         280         290         300
CCTCCTTCTT CTCGCTCTAC ACCTTCTGCT CCTCCTCCAT CTCCACCAAC TCCATCTACG
       310         320         330         340         350         360
CCGGGATCTC CACCTCCTCT TCCTCAGCCG TCTCCACCCG CTCCAACTAC GCCCGGATCT
       370         380         390         400         410         420
CCACCCGCAC CTGTTACTCC TCCTACTCGA AACCCTCCAC CTTCAGTCCC AGGACCACCG
       430         440         450         460         470         480
TCCAATCCTT CACGCGAAGG AGGATCTCCT CGACCTCCAT CTTCTCCCTC GCCGCCGTCT
       490         500         510         520         530         540
CCTTCTTCCG ACGGTTTATC AACAGGAGTG GTGGTGGGAA TCGCCATCGG AGGAGTCGCT
       550         560         570         580         590         600
CTGCTTGTGA TAGTGACTCT GATTTGTCTC CTCTGTAAGA AGAAACGACG GAGAGACGAA
       610         620         630         640         650         660
GAAGATGCTT ACTATGTTCC TCCGCCACCT CCTCCTGGTC CCAAAGCCGG AGGACCTTAC
       670         680         690         700         710         720
GGTGGACAGC AGCAACAATG GCGGCAACAA AACGCAACAC CACCGTCAGA TCATGTCGTG
       730         740         750         760         770         780
ACGTCACTAC CACCACCACC TAAGGCTCCA TCTCCACCAC GGCAACCTCC TCCACCTCCA
       790         800         810         820         830         840
CCACCGCCTT TCATGAGCAG CAGCGGCGGC TCCGACTACT CGGACCGTCC AGTTCTTCCT
       850         860         870         880         890         900
CCACCGTCTC CAGGGCTTGT GTTAGGCTTC TCCAAAAGCA CTTTCACATA CGAGGAGCTA
       910         920         930         940         950         960
GCTAGAGCCA CCAATGGTTT CTCCGAGGCG AACTTGTTAG GACAAGGCGG GTTCGGTTAC
       970         980         990        1000        1010        1020
GTGCACAAAG GTGTGTTGCC TAGTGGGAAA GAAGTTGCTG TGAAGCAGTT GAAAGTTGGG
      1030        1040        1050        1060        1070        1080
AGTGGTCAGG GAGAGAGGGA GTTTCAGGCA GAGGTTGAGA TCATCAGCAG AGTTCACCAC
      1090        1100        1110        1120        1130        1140
AGGCATCTGG TGTCTCTTGT TGGTTATTGC ATCGCCGGTG CCAAAAGATT GCTTGTCTAT
      1150        1160        1170        1180        1190        1200
GAGTTTGTTC CTAACAACAA TCTCGAGCTT CACCTCCATG GCGAGGGACG GCCTACAATG
      1210        1220        1230        1240        1250        1260
GAATGGAGCA CCAGATTGAA GATTGCTCTT GGATCTGCTA AAGGACTTTC TTATCTTCAT
      1270        1280        1290        1300        1310        1320
GAAGATTGCA ATCCTAAAAT CATTCACCGT GATATCAAGG CTTCAAACAT ATTGATAGAT
      1330        1340        1350        1360        1370        1380
TTCAAGTTTG AAGCTAAGGT TGCTGATTTT GGTCTTGCTA AGATTGCTTC TGATACAAAC
      1390        1400        1410        1420        1430        1440
ACGCATGTAT CAACACGTGT GATGGGAACC TTTGGGTACT TGGCTCCGGA ATACGCTGCA
      1450        1460        1470        1480        1490        1500
AGCGGAAAGC TCACGGAGAA GTCTGACGTT TTCTCATTTG GCGTTGTGCT TTTGGAGCTC
      1510        1520        1530        1540        1550        1560
```

Figure 1(d) (continued on next page)

```
ATTACTGGAC GTCGACCCGT TGATGCCAAC AATGTCTATG TAGATGACAG CTTAGTTGAC
      1570        1580        1590        1600        1610        1620
TGGGCACGAC CATTGCTTAA CCGAGCATCT GAGCAAGGAG ACTTTGAGGG TTTAGCTGAT
      1630        1640        1650        1660        1670        1680
GCAAAGATGA ATAATGGGTA TGACAGAGAG GAGATGGCTC GCATGGTTGC TTGTGCTGCG
      1690        1700        1710        1720        1730        1740
GCTTGTGTTC GCCATTCAGC TCGCCGCAGA CCTCGCATGA GCCAGATTGT GCGTGCGTTA
      1750        1760        1770        1780        1790        1800
GAAGGAAATG TATCACTGTC AGATCTTAAC GAAGGGATGA GACCAGGTCA AAGCAATGTA
      1810        1820        1830        1840        1850        1860
TACAGCTCAT ACGGAGGAAG CACCGATTAT GACTCGAGCC AGTACAATGA AGACATGAAG
      1870        1880        1890        1900        1910        1920
AAGTTTAGGA AAATGGCACT TGGAACTCAA GAGTACAACG CCACGGGTGA GTACAGTAAT
      1930        1940        1950        1960        1970        1980
CCGACCAGTG ACTATGGACT GTACCCGTCT GGTTCAAGCA GCGAGGGCCA AACCACACGC
      1990        2000        2010        2020        2030        2040
GAAATGGAGA TGGGGAAGAT TAAGAGAACC GGTCAGGGTT ATAGTGGACC TTCTCTTTAA
      2050        2060        2070        2080        2090        2100
ACCAGATGGG AGAGAAATTG AAGGGTGTTT TTTCATTATT TTTTTAAAAC TGTAAAGATA
      2110        2120        2130        2140        2150        2160
TGAGAAAATT GCCTTACTCT AATTAAAACC ACTACGATAT AAGGTTATAA TACGTTTTGA
      2170        2180        2190        2200        2210        2220
ATTGGTTTTT AAAAAAAAAA AAAAAAAAA. .......... .......... ..........
```

Figure 1(d) (continued)

```
LTLWSPCLLS SPAASF*HSL HLPF*FRSKE ADMSSAPSPG TGSPPSPPSN STTTTPPPAS
APPPTTPSSP PPPSTIPTSP PPSSRSTPSA PPPSPPTPST PGSPPPLPQP SPPAPTTPGS
PPAPVTPPTR NPPPSVPGPP SNPSREGGSP RPPSSPSPPS PSSDGLSTGV VVGIAIGGVA
LLVIVTLICL LCKKKRRRDE EDAYYVPPPP PPGPKAGGPY GGQQQQWRQQ NATPPSDHVV
TSLPPPPKAP SPPRQPPPPP PPPFMSSSGG SDYSDRPVLP PPSPGLVLGF SKSTFTYEEL
ARATNGFSEA NLLGQGGFGY VHKGVLPSGK EVAVKQLKVG SGQGEREFQA EVEIISRVHH
RHLVSLVGYC IAGAKRLLVY EFVPNNNLEL HLHGEGRPTM EWSTRLKIAL GSAKGLSYLH
EDCNPKIIHR DIKASNILID FKFEAKVADF GLAKIASDTN THVSTRVMGT FGYLAPEYAA
SGKLTEKSDV FSFGVVLLEL ITGRRPVDAN NVYVDDSLVD WARPLLNRAS EQGDFEGLAD
AKMNNGYDRE EMARMVACAA ACVRHSARRR PRMSQIVRAL EGNVSLSDLN EGMRPGQSNV
YSSYGGSTDY DSSQYNEDMK KFRKMALGTQ EYNATGEYSN PTSDYGLYPS GSSSEGQTTR
EMEMGKIKRT GQGYSGPSL- TRWERN*RVF FHYFFKTVKI *ENCLTLIKT TTI*GYNTF*
IGF*KKKKK
```

Figure 1(e)

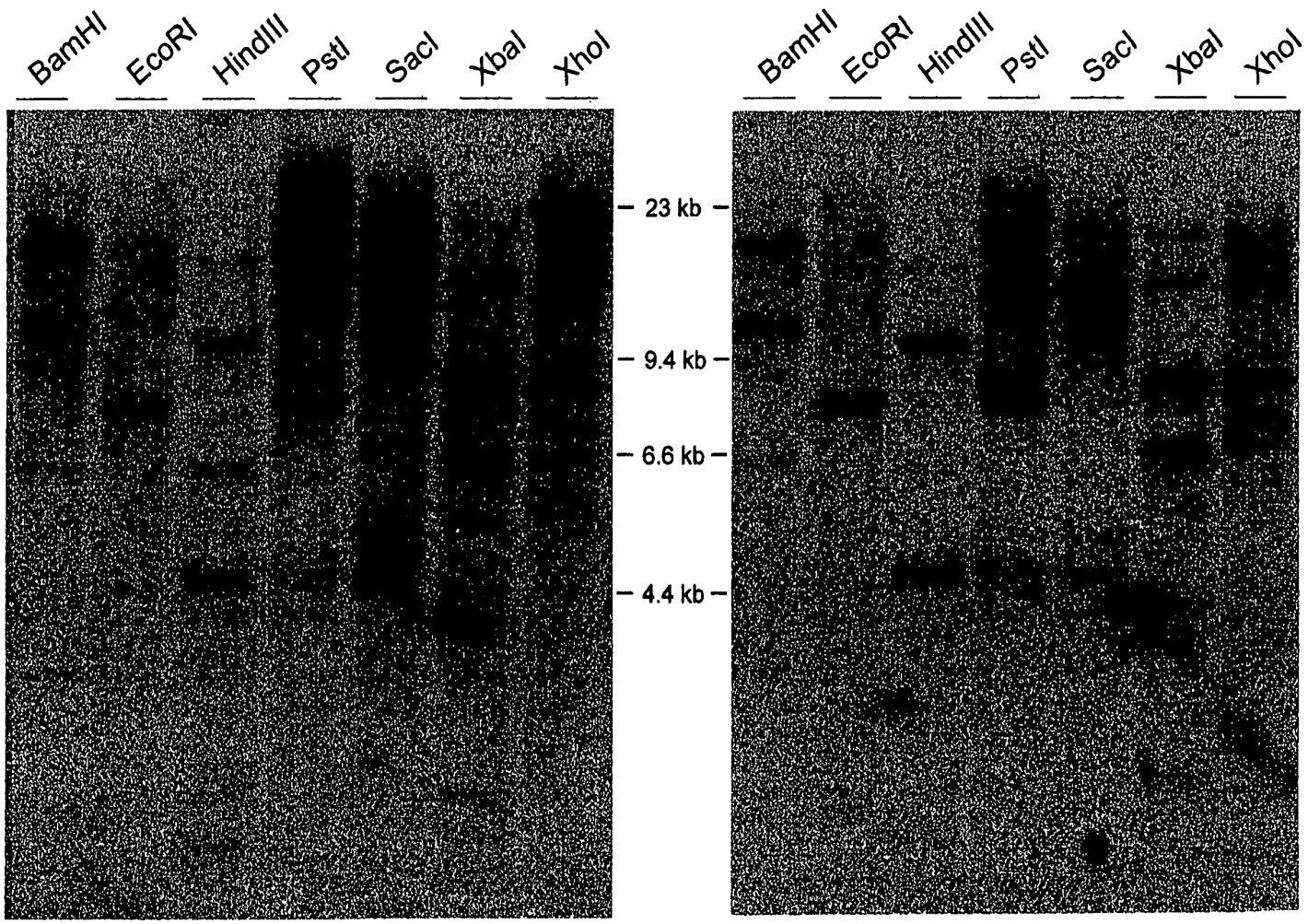


Figure 2

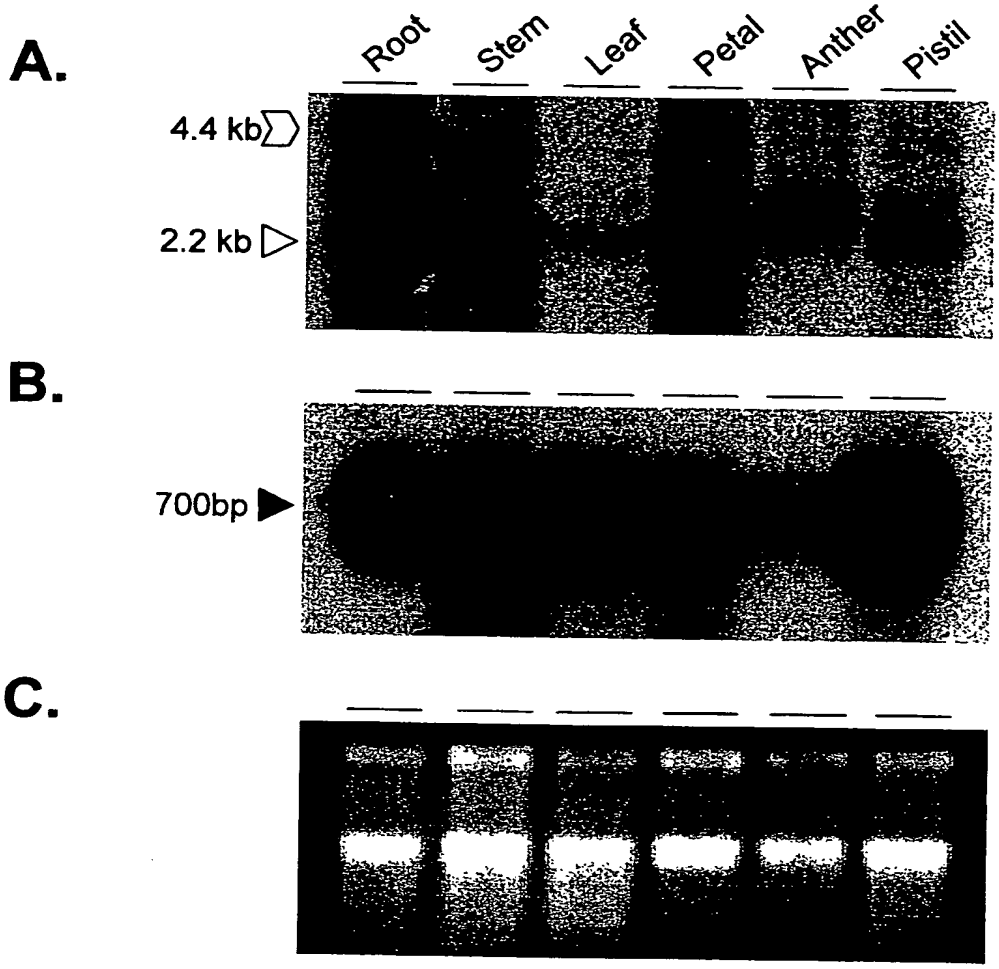


Figure 3

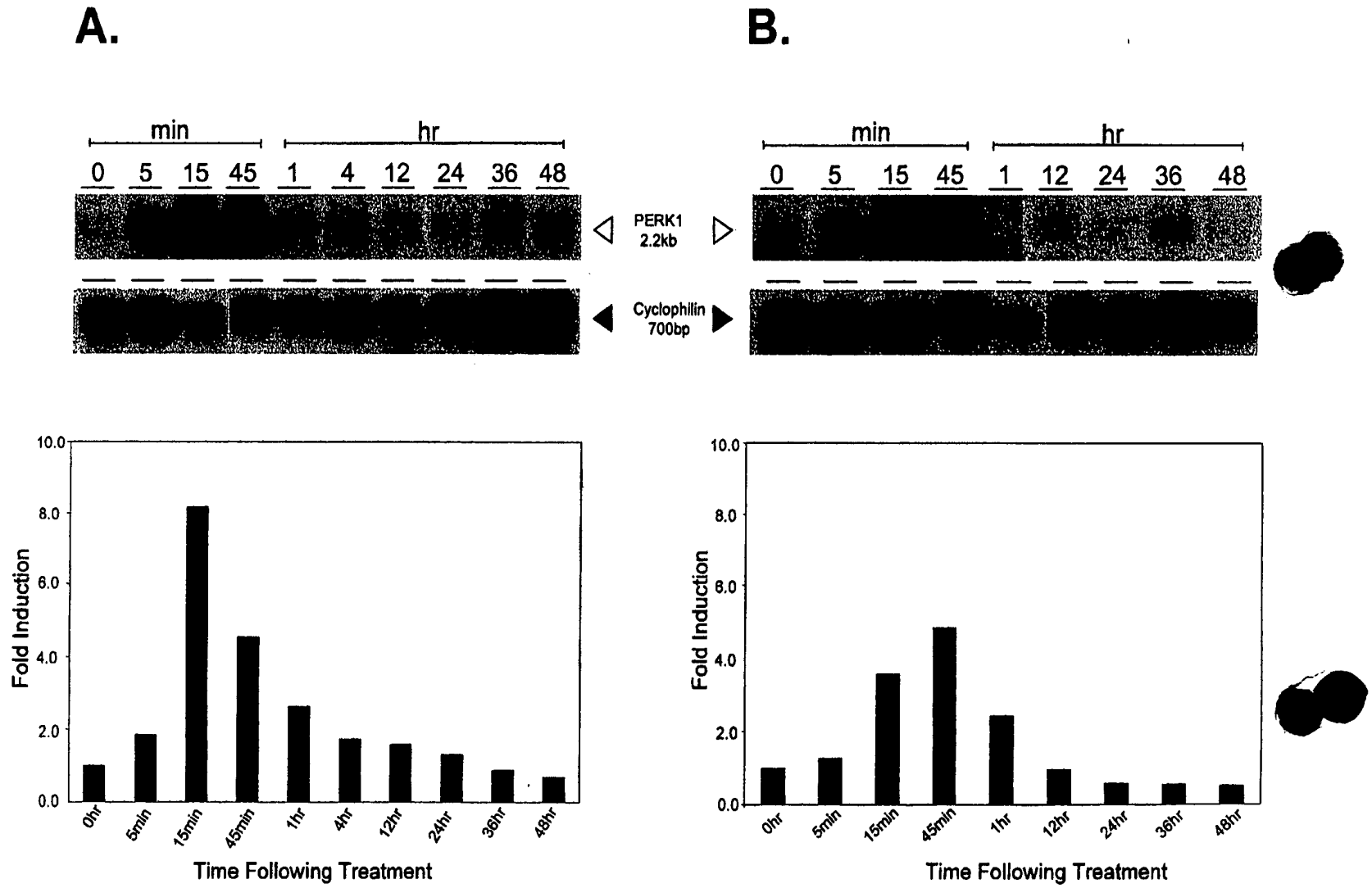


Figure 4

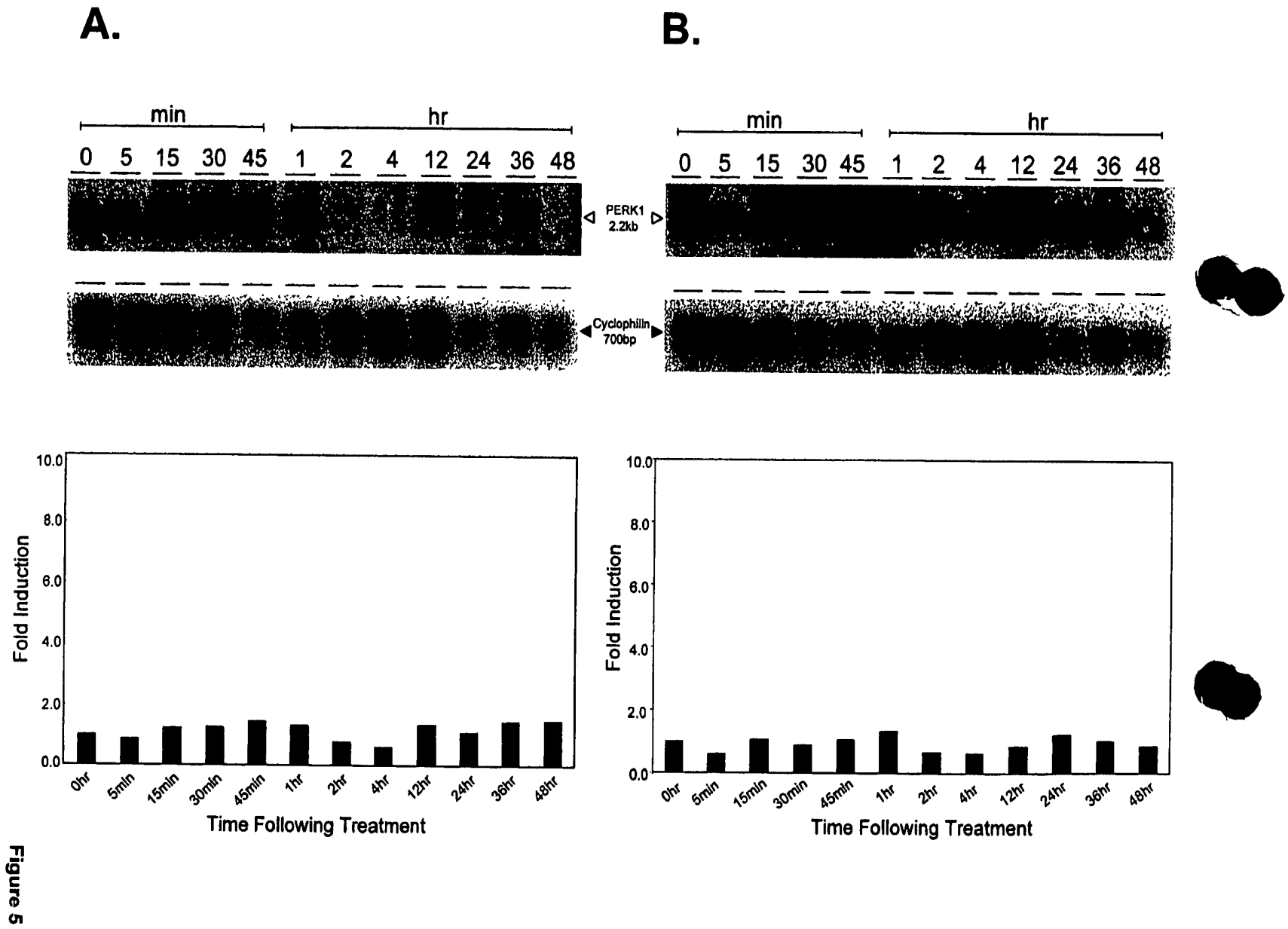


Figure 5

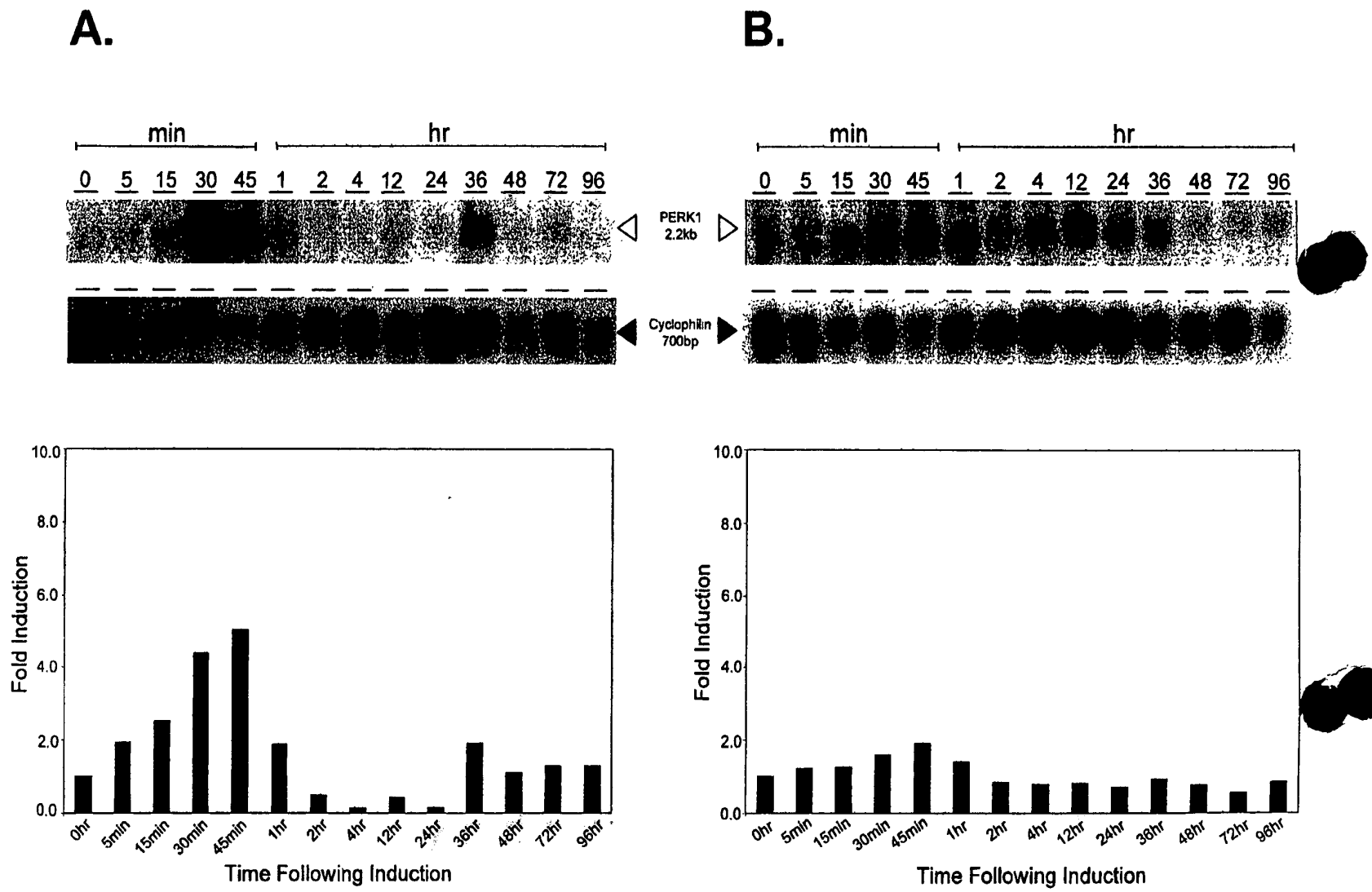


Figure 6

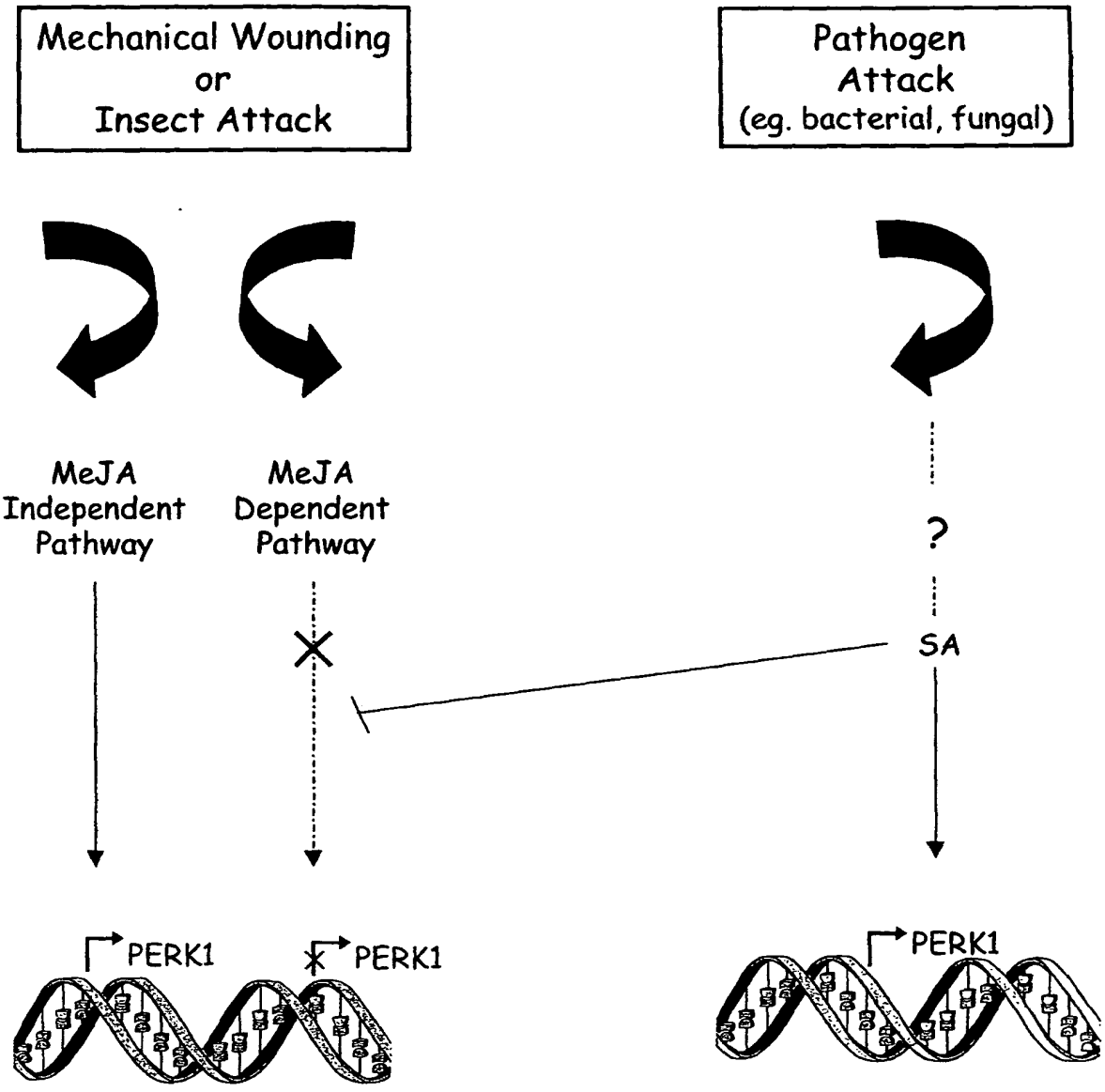

Mechanical Wounding
or
Insect Attack
Pathogen
Attack
(eg. bacterial, fungal)
MeJA
Independent
Pathway
MeJA
Dependent
Pathway
?
SA
PERK1
PERK1
PERK1

Figure 7

Bacterially Expressed Extracellular Domain of PERK1

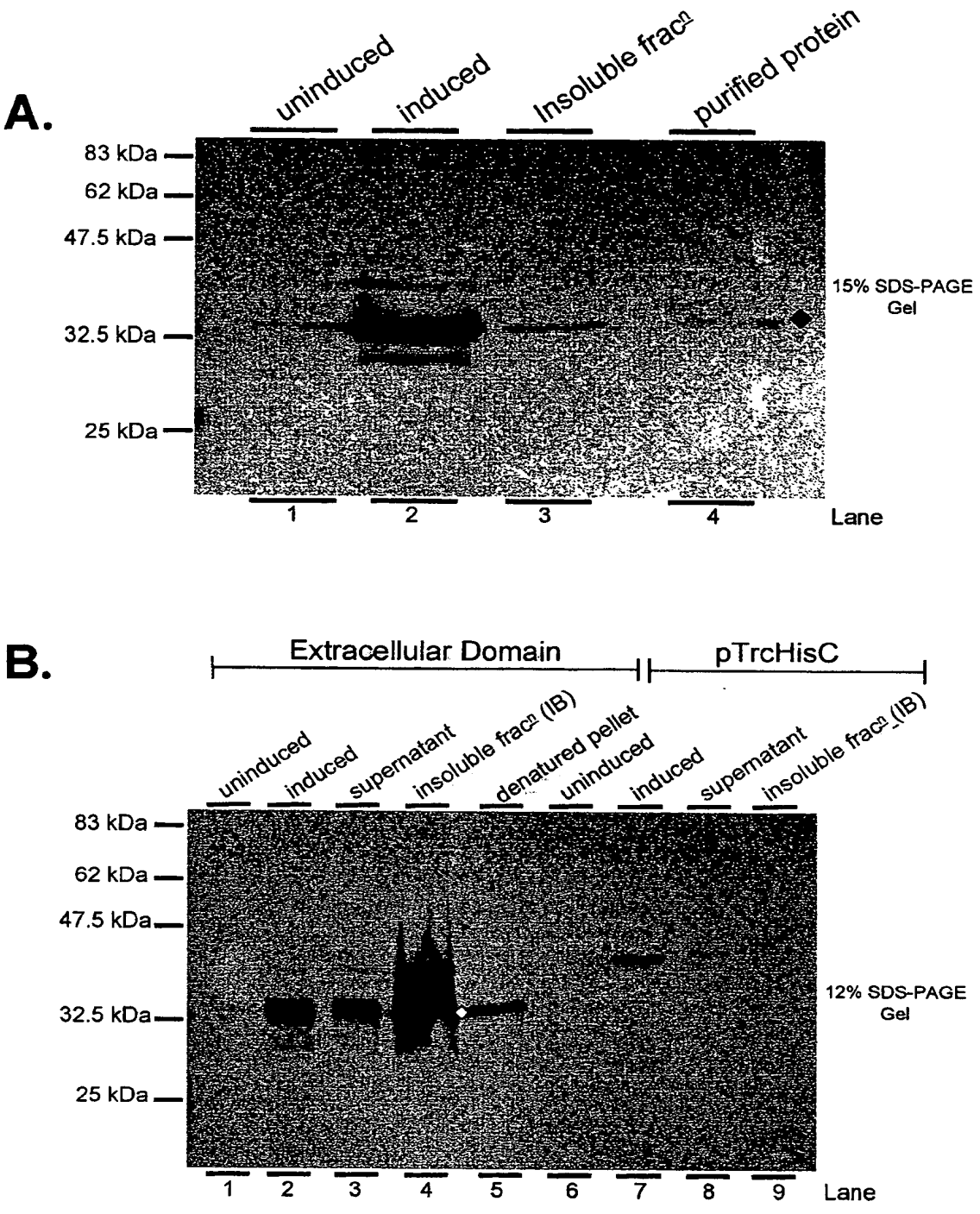


Figure 8

Bacterially [redacted]pressed Full Lengt[redacted]PERK1

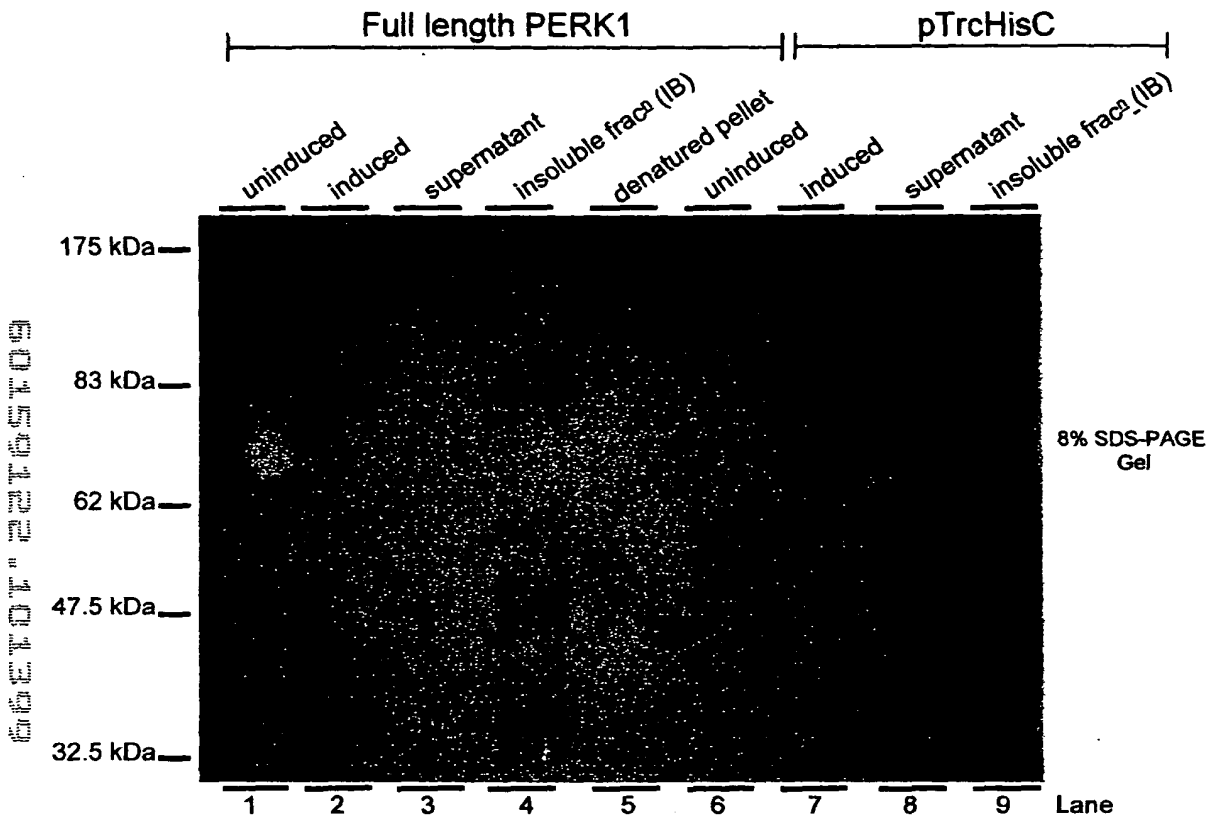
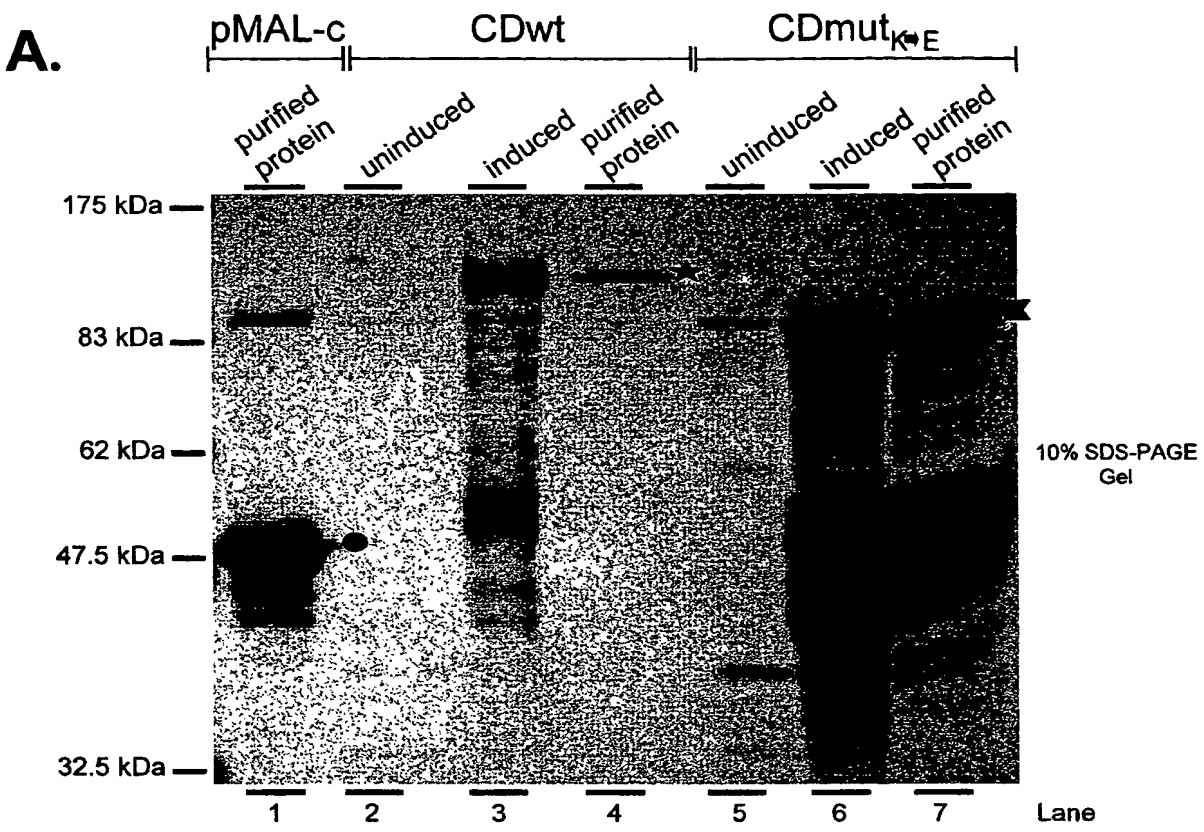


Figure 9

Bacterial Expression and Kina e Assay of Wild-Typ nd Mutated
atalytic Domains of PERK1



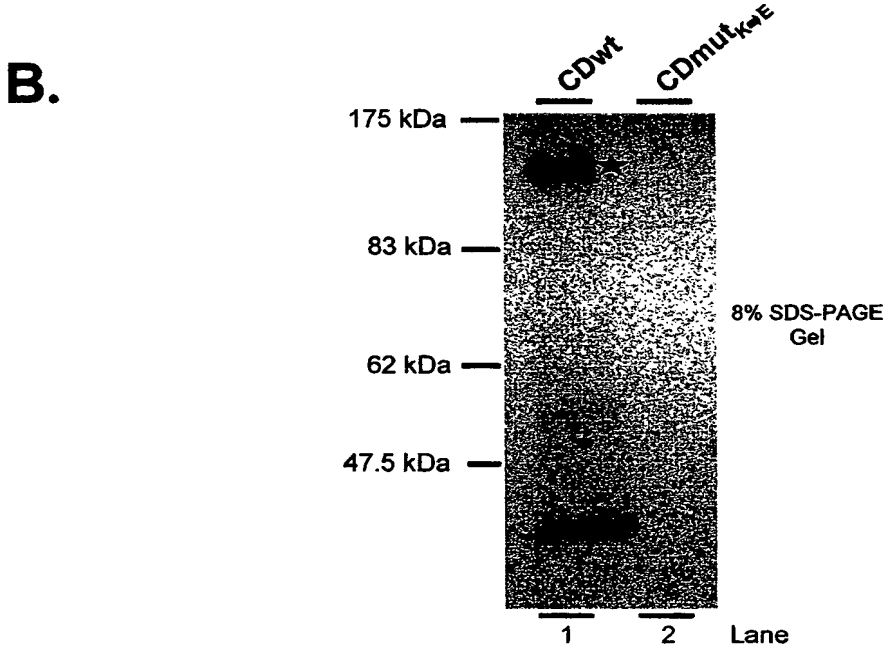
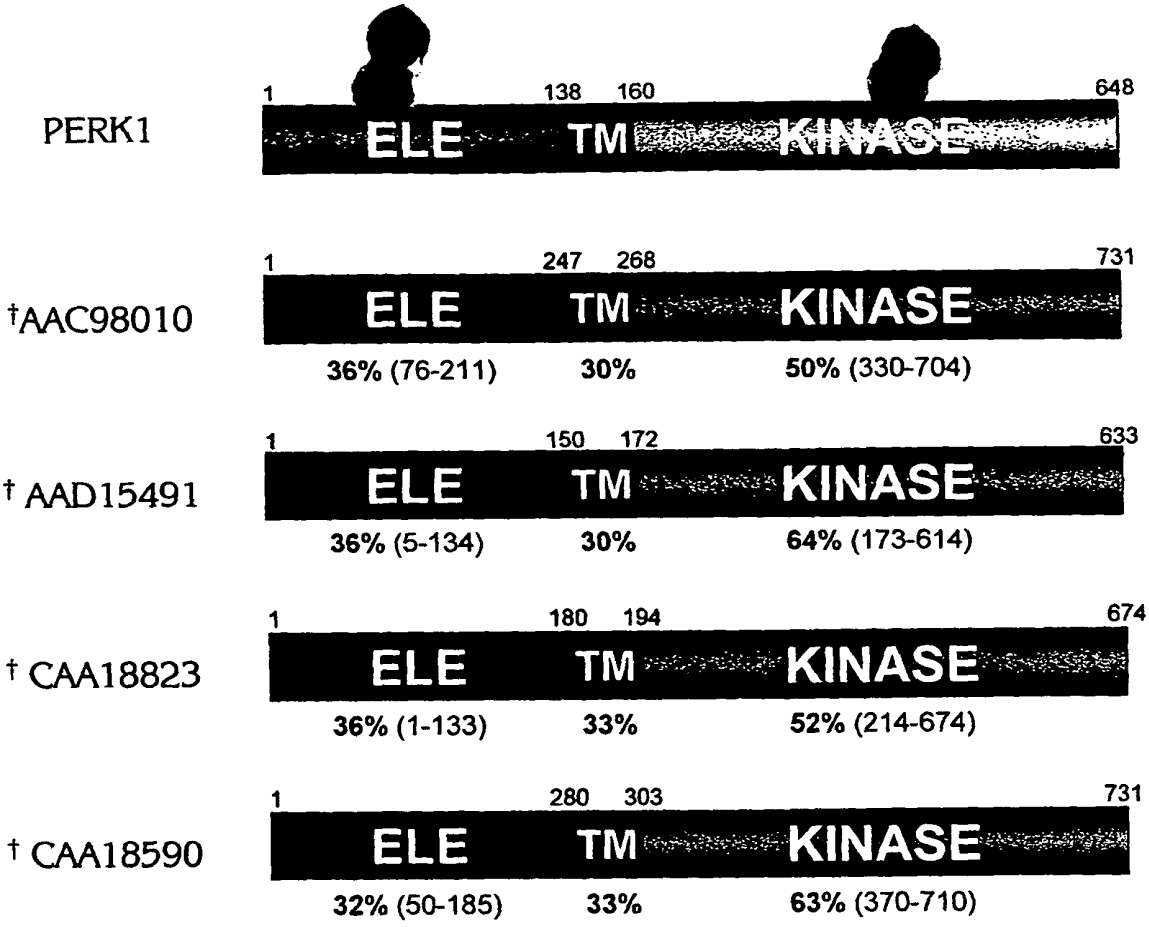


Figure 10



All have no signal peptide and predicted to be Type1b integral membrane proteins

† Predicted proteins from the Arabidopsis genome sequencing project.

ELE: Extensin-like extracellular ; TM: Transmembrane domain ; Red = sequence identity to PERK1 domain

	PERK1	ACC98010	AAD15491	CAA18823	CAA18590
PERK1		50%	64%	52%	63%
ACC98010	36%		54%	50%	52%
AAD15491	36%	25%		63%	51%
CAA18823	36%	29%	27%		60%
CAA18590	32%	36%	30%	27%	

Kinase Domains

ELE Domains

Figure 11

Figure 12 A.

TAGAAAAAAA AAAATGTCAG ACTTAGGCGA GTCGCCGAGT TCTTCACCAC CAGCACCACC
AGCTGATACC GCTCCTCCAC CAGAGACTCC ATCAGAAAAC TCAGCTCTTC CACCTGTTGA
TTCCTCTCCT CCTAGTCCAC CAGCTGATTC ATCATCAACA CCGCCGCTGT CAGAACCATC
CACTCCTCCT CCAGATTCAC AGCTTCCTCC TTTACCTTCG ATTCTTCCTC CGCTAACAGA
TTCTCCACCT CCACCTTCCG ATTCTTCTCC ACCCGTTGAT TCAACCCCTT CTCCGCCGCC
ACCGACGTCA AACGAATCTC CTTCTCCTCC AGAAGATTCC GAAACACCAC CTGCTCCACC
AAATGAATCC AATGACAACA ACCCTCCTCC GTCTCAAGAT CTTCAATCGC CTCCTCCATC
GTCGCCGTCG CCGAATGTAG GACCCACAAA CCCGGAATCA CCACCGTTAC AATCTCCTCC
AGCTCCACCA GCATCAGATC CTACAAATTC ACCGCCAGCT TCACCATTAG ACCCTACCAA
TCCTCCCCCA ATACAACCAT CAGGACCAGC CACTTCTCCT CCGGCTAATC CCAACGCTCC
GCCGAGCCCA TTCCCCACAG TACCACCCAA AACTCCTTCT AGTGGACCTG TGGTGTCTCC
ATCTCTCACA TCCCCTAGTA AAGGAACTCC TACTCCAAAC CAAGGCAATG GAGATGGCGG
TGGCGGTGGT GGCGGCTATC AAGGGAAGAC TATGGTTGGT ATGGCTGTAG CCGGTTTCGC
AATCATGGCG CTTATAGGCG TTGTGTTCTT AGTGAGAAGA AAGAAAAAGA GAAACATTGA
TAGCTATAAT CACTCACAGT ACTTGCCACA TCCCAATTTC TCTGTTAAAT CAGGTTTAAA
AATCTCACCT TTATCTCTCT CTGATCATCT TCTATGTGCT TGAATCATCT CTCTGACTAT
CTTTGCTTTT GATGTAGATG GATTCTTATA CGGTCAAGAT CCAGGTAAAG GATACTCCTC
TGGTCCTAAT GGTTCAATGT ATAACAATTC ACAGCAACAA CAATCCTCTA TGGGAAACAG
TTATGGTACA GCTGGTGGTG GTTATCCTCA TCATCAAATG CAATCAAGTG GCACACCTGA
CTCTGCTATA CTCGGAAGTG GCCAGACTCA TTTCAGTTAC GAAGAGCTTG CTGAGATAAC
ACAAGGCTTT GCTCGCAAAA ACATTCTTGG AGAAGGCGGA TTTGGATGTG TCTATAAAGG
TACATTGCAG GATGGTAAAG TTGTTGCGGT TAAGCAGCTT AAAGCTGGAA GTGGACAAGG
TGACCGTGAA TTCAAAGCAG AGGTTGAGAT CATCAGCCGC GTTCATCATC GCCATTTGGT
CTCTCTGGTT GGTTACTGCA TTTCAGACCA GCATAGATTG CTTATCTATG AGTATGTTTC
TAATCAAACC TTGGAGCATC ATTTGCATGG TGAGTGACTT GTTACCATTT TCGTTATAGA
TAAGACTTTT TTTTAGCTTT ACGTGTTAGA CTGACTCGCT TTACGCTTTA GGAAAGGGTT
TGCCAGTTTT AGAGTGGTCT AAGAGAGTCC GGATCGCTAT AGGATCAGCC AAAGGGTTGG
CATATCTTCA CGAAGACTGT AAGTAATGCC TTCACATTTT CTTAGTTGTG TGCTTTGGTT
ATGCACTTCA TAGTTTAAAC AGAAGCCAAA AATCATATCC TTGTTTTATT TTACAGGTCA
TCCGAAAATC ATTCACAGAG ATATAAAGTC AGCAAATATT CTTCTAGATG ATGAATATGA
AGCTCAGGCA ATAATGAAAT CCTCCTTTTC GTTAAATCTA TCTTATGACT GTAAAGTTTT
AGTTAATGAG ACTTGTTCTG TTTTTTGGAT GTTTAGGTTG CTGATTTTGG ACTTGCTAGA
CTCAATGATA CAACACAAAC TCATGTTTCA ACTCGGGTTA TGGGAACCTT CGGGTAAGCA
AACATTCATC ACAAACTCTA CTCCAAAACT GGACCTTATT GATCCAATGC CTGATGAAAA
GTTTGTTATA TATGGCTTGA GGCAACAAAT TGGATCAAAC CTGAATCTTT ATTGATCGTA
TGGCTGCATG ACATGTTTTG TGTTAAGGTA CCTAGCGCCG GAATATGCAT CAAGTGGAAA
ATTGACTGAT AGATCCGATG TATTCTCATT CGGGGTTGTT CTCTTAGAGC TTGTAACTGG
ACGGAAACCA GTTGACCAGA CTCAGCCTCT AGGAGAAGAG AGTTTGGTTG AATGGGTAAG
AATCCAACTT TCAAACATTC TTCAATAATA GTAAGATTGG CCCTAGTATA CTTATATAGT
ACTTATAAAT GAACTCACAG GCGCGCCCGC TGCTTCTCAA AGCCATTGAG ACCGGAGATT
TAAGCGAACT GATTGATACA CGGCTTGAAA AGCGTTATGT GGAGCATGAA GTCTTCAGAA
TGATCGAGAC AGCCGCTGCA TGTGTTAGAC ATTCTGGTCC AAAACGCCCA CGCATGGTTC
AGGTAATTCT GACTAACCAA AAGTCCAAAG CTCCCATATA TAGTAACAAG TGATTTCTCA
CATCTGAAAA CTTATCTACT CTTCGAAATA AGGTTGTGAG AGCATTGGAC TGCGACGGAG
ACTCGGGAGA TATTAGCAAC GGAATCAAAA TTGGGCAAAG CACAACTTAT GACTCAGGGC
AATACAATGA AGACATTATG AAATTCAGGA AAATGGCGTT TGGTGGTGAT AACAGCGTAG
AGTCAGGATT GTACAGTGGA AACTACTCTG CCAAAAGCTC TTCAGATTTC TCAGGGAATG
AATCTGAGAC TCGGCCTTTC AACAACCGAC GGTTCTGATC ATACAATAGG TGAAAGTAAC

Figure 12 B.

```
   1  M  S  D  L  G  E  S  P  S  S  S  P  P  A  P  P  A  D  T  A      20
   1 ATG TCA GAC TTA GGC GAG TCG CCG AGT TCT TCA CCA CCA GCA CCA CCA GCT GAT ACC GCT      60
  21  P  P  P  E  T  P  S  E  N  S  A  L  P  P  V  D  S  S  F  P      40
  61 CCT CCA CCA GAG ACT CCA TCA GAA AAC TCA GCT CTT CCA CCT GTT GAT TCC TCT CCT CCT     120
  41  S  P  P  A  D  S  S  S  T  P  P  L  S  E  P  S  T  P  P  P      60
 121 AGT CCA CCA GCT GAT TCA TCA TCA ACA CCG CCG CTG TCA GAA CCA TCC ACT CCT CCT CCA     180
  61  D  S  Q  L  P  P  L  P  S  I  L  P  P  L  T  D  S  P  P  P      80
 181 GAT TCA CAG CTT CCT CCT TTA CCT TCG ATT CTT CCT CCG CTA ACA GAT TCT CCA CCT CCA     240
  81  P  S  D  S  S  P  F  V  D  S  T  P  S  P  P  P  P  T  S  N     100
 241 CCT TCC GAT TCT TCT CCA CCC GTT GAT TCA ACC CCT TCT CCG CCG CCA CCG ACG TCA AAC     300
 101  E  S  P  S  P  P  E  D  S  E  T  P  P  A  P  P  N  E  S  N     120
 301 GAA TCT CCT TCT CCT CCA GAA GAT TCC GAA ACA CCA CCT GCT CCA CCA AAT GAA TCC AAT     360
 121  D  N  N  P  P  P  S  Q  D  L  Q  S  P  P  P  S  S  P  S  P     140
 361 GAC AAC AAC CCT CCT CCG TCT CAA GAT CTT CAA TCG CCT CCT CCA TCG TCG CCG TCG CCG     420
 141  N  V  G  P  T  N  P  E  S  P  P  L  Q  S  P  P  A  P  P  A     160
 421 AAT GTA GGA CCC ACA AAC CCG GAA TCA CCA CCG TTA CAA TCT CCT CCA GCT CCA CCA GCA     480
 161  S  D  P  T  N  S  P  P  A  S  P  L  D  P  T  N  P  P  P  I     180
 481 TCA GAT CCT ACA AAT TCA CCG CCA GCT TCA CCA TTA GAC CCT ACC AAT CCT CCC CCA ATA     540
 181  Q  P  S  G  P  A  T  S  P  P  A  N  P  E  A  P  P  S  P  F     200
 541 CAA CCA TCA GGA CCA GCC ACT TCT CCT CCG GCT AAT CCC AAC GCT CCG CCG AGC CCA TTC     600
 201  P  T  V  P  P  K  T  P  S  S  G  P  V  V  S  P  S  L  T  S     220
 601 CCC ACA GTA CCA CCC AAA ACT CCT TCT AGT GGA CCT GTG GTG TCT CCA TCT CTC ACA TCC     660
 221  P  S  K  G  T  P  T  P  N  Q  G  N  G  D  G  G  G  G  G  G     240
 661 CCT AGT AAA GGA ACT CCT ACT CCA AAC CAA GGC AAT GGA GAT GGC GGT GGC GGT GGT GGC     720
 241  G  Y  Q  G  K  T  M  V  G  M  A  V  A  G  F  A  I  M  A  L     260
 721 GGC TAT CAA GGG AAG ACT ATG GTT GGT ATG GCT GTA GCC GGT TTC GCA ATC ATG GCG CTT     780
 261  I  G  V  V  F  L  V  R  R  K  K  K  R  N  I  D  S  Y  N  H     280
 781 ATA GGC GTT GTG TTC TTA GTG AGA AGA AAG AAA AAG AGA AAC ATT GAT AGC TAT AAT CAC     840
 281  S  Q  Y  L  P  H  P  N  F  S  V  K  S  D  G  F  L  Y  G  Q     300
 841 TCA CAG TAC TTG CCA CAT CCC AAT TTC TCT GTT AAA TCA GAT GGA TTC TTA TAC GGT CAA     900
 301  D  P  G  K  G  Y  S  S  G  P  N  G  S  M  Y  N  N  S  Q  Q     320
 901 GAT CCA GGT AAA GGA TAC TCC TCT GGT CCT AAT GGT TCA ATG TAT AAC AAT TCA CAG CAA     960
 321  Q  Q  S  S  M  G  N  S  Y  G  T  A  G  G  G  Y  P  H  H  Q     340
 961 CAA CAA TCC TCT ATG GGA AAC AGT TAT GGT ACA GCT GGT GGT GGT TAT CCT CAT CAT CAA    1020
 341  M  Q  S  S  G  T  P  D  S  A  I  L  G  S  G  Q  T  H  F  S     360
1021 ATG CAA TCA AGT GGC ACA CCT GAC TCT GCT ATA CTC GGA AGT GGC CAG ACT CAT TTC AGT    1080
 361  Y  E  E  L  A  E  I  T  Q  G  F  A  R  K  N  I  L  G  E  G     380
1081 TAC GAA GAG CTT GCT GAG ATA ACA CAA GGC TTT GCT CGC AAA AAC ATT CTT GGA GAA GGC    1140
 381  G  F  G  C  V  Y  K  G  T  L  Q  D  G  K  V  V  A  V  K  Q     400
1141 GGA TTT GGA TGT GTC TAT AAA GGT ACA TTG CAG GAT GGT AAA GTT GTT GCG GTT AAG CAG    1200
 401  L  K  A  G  S  G  Q  G  D  R  E  P  K  A  E  V  E  I  I  S     420
1201 CTT AAA GCT GGA AGT GGA CAA GGT GAC CGT GAA TTC AAA GCA GAG GTT GAG ATC ATC AGC    1260
 421  R  V  H  H  R  H  L  V  S  L  V  G  Y  C  I  S  D  Q  H  R     440
1261 CGC GTT CAT CAT CGC CAT TTG GTC TCT CTG GTT GGT TAC TGC ATT TCA GAC CAG CAT AGA    1320
 441  L  L  I  Y  E  Y  V  S  N  Q  T  L  E  H  H  L  H  E  W  S     460
1321 TTG CTT ATC TAT GAG TAT GTT TCT AAT CAA ACC TTG GAG CAT CAT TTG CAT GAG TGG TCT    1380
 461  K  R  V  R  I  A  I  G  S  A  K  G  L  A  Y  L  H  E  D  C     480
1381 AAG AGA GTC CGG ATC GCT ATA GGA TCA GCC AAA GGG TTG GCA TAT CTT CAC GAA GAC TGT    1440
 481  H  P  K  I  I  H  R  D  I  K  S  A  N  I  L  L  D  D  E  Y     500
1441 CAT CCG AAA ATC ATT CAC AGA GAT ATA AAG TCA GCA AAT ATT CTT CTA GAT GAT GAA TAT    1500
 501  E  A  Q  A  I  M  K  S  S  F  S  L  N  L  S  Y  D  C  K  V     520
1501 GAA GCT CAG GCA ATA ATG AAA TCC TCC TTT TCG TTA AAT CTA TCT TAT GAC TGT AAA GTT    1560
 521  L  V  A  D  F  G  L  A  R  L  N  D  T  T  Q  T  H  V  S  T     540
1561 TTA GTT GCT GAT TTT GGA CTT GCT AGA CTC AAT GAT ACA ACA CAA ACT CAT GTT TCA ACT    1620
 541  R  V  M  G  T  F  G  Y  L  A  P  E  Y  A  S  S  G  K  L  T     560
1621 CGG GTT ATG GGA ACC TTC GGG TAC CTA GCG CCG GAA TAT GCA TCA AGT GGA AAA TTG ACT    1680
 561  D  R  S  D  V  F  S  F  G  V  V  L  L  E  L  V  T  G  R  K     580
1681 GAT AGA TCC GAT GTA TTC TCA TTC GGG GTT GTT CTC TTA GAG CTT GTA ACT GGA CGG AAA    1740
 581  P  V  D  Q  T  Q  F  L  G  E  E  S  L  V  E  W  A  R  P  L     600
1741 CCA GTT GAC CAG ACT CAG CCT CTA GGA GAA GAG AGT TTG GTT GAA TGG GCG CGC CCG CTG    1800
 601  L  L  K  A  I  E  T  G  D  L  S  S  L  I  D  T  R  L  H  K     620
1801 CTT CTC AAA GCC ATT GAG ACC GGA GAT TTA AGC GAA CTG ATT GAT ACA CGG CTT CAA AAG    1860
 621  R  Y  V  E  H  E  V  F  R  M  I  E  T  A  A  A  C  V  R  H     640
1861 CGT TAT GTG GAG CAT GAA GTC TTC AGA ATG ATC GAG ACA GCC GCT GCA TGT GTT AGA CAT    1920
 641  S  G  P  K  R  P  R  M  V  Q  V  V  R  A  L  D  C  D  G  D     660
1921 TCT GGT CCA AAA CGC CCA CGC ATG GTT CAG GTT GTG AGA GCA TTG GAC TGC GAC GGA GAC    1980
 661  S  G  D  I  S  N  G  I  K  I  G  Q  S  T  T  Y  D  S  G  Q     680
1981 TCG GGA GAT ATT AGC AAC GGA ATC AAA ATT GGG CAA AGC ACA ACT TAT GAC TCA GGG CAA    2040
 681  Y  N  E  D  I  M  K  F  R  K  M  A  P  G  G  D  N  S  V  E     700
2041 TAC AAT GAA GAC ATT ATG AAA TTC AGG AAA ATG GCG TTT GGT GGT GAT AAC AGC GTA GAG    2100
 701  S  G  L  Y  S  G  N  Y  S  A  K  S  S  S  D  F  S  G  N  E     720
2101 TCA GGA TTG TAC AGT GGA AAC TAC TCT GCC AAA AGC TCT TCA GAT TTC TCA GGG AAT GAA    2160
 721  S  E  T  R  P  F  N  N  R  R  F  *                             732
2161 TCT GAG ACT CGG CCT TTC AAC AAC CGA CGG TTC TGA                    2196
```

Figure 13 A.

```
GAAAATTTTG ATCTCCGATG GCTTCTTCTC CTGAATCTGC TCCTCCAACA AACTCCACCT
CTTCTCCATC TCCACCGTCT AATACCAATT CAACCACCTC TTCTCCGCCG GCTCCGTCTC
CTCCTTCTCC TACACCTCCT CAAGGAGACT CATCATCATC GCCACCTCCT GATTCCACAT
CTCCACCAGC TCCACAAGCT CCTAACCCTC CTAATTCCTC TAATAACTCT CCTTCCCCTC
CGTCACAGGG CGGTGGAGGA GAAAGAGGAA ATGGAGGAAA CAATGGTGGC AATGATACTC
CACCGTCACG CGGCTCTCCT CCTTCTCCTC CTTCTAGGAG TAATGGAGAT AATGGTGGTA
GCAGATCATC GCCACCAGGA GACACTGGAG GCTCTCGCTC AGACAACCCT CCTTCTAGCG
GAGGAAGCAG TGGAGGAGGT GGAGGTGGAA GAAGTAATAC GAATACAGCG ATCATAGTTG
GTGTATTAGT CGGAGCTGGA CTTTTGATGA TCGTTCTTAT TATTGTGTGT CTTAGACGCA
AAAAGAAGAG AAAAGACTCC TTCTACCCTG AACCCATGAA AGGTAAAAAC ATATACACAC
TCTTATGTTT CAACAAATAA GAAGCTTAGA TTCTTTCATA AAATTTCAGG AAACCAATAT
CAATACTATG GAAACAACAA CAACAACAAT GCTTCACAGA ATTATCCGAA TTGGCACCTA
AATTCACAAG GCCAAAACCA ACAATCTACT GGTGGTTGGG GAGGCGGTGG ACCATCACCG
CCTCCTCCTC CGCGGATGCC TACAAGCGGA GAAGATTCTT CCATGTACTC AGGCCCATCA
CGCCCAGTTT TACCTCCTCC TTCGCCTGCT CTAGCCCTCG GATTCAACAA GAGCACTTTT
ACTTACCAAG AGCTTGCGGC TGCAACAGGA GGGTTTACGG ATGCTAACCT TTTGGGACAG
GGAGGATTTG GGTATGTCCA TAAAGGAGTC TTGCCTAGCG GGAAAGAAGT AGCAGTTAAG
AGTTTAAAAG CGGGTAGCGG ACAAGGAGAG AGGGAGTTTC AAGCTGAGGT CGATATCATT
AGCCGTGTGC ATCATCGGTA TCTTGTTTCT TTGGTTGGAT ATTGCATAGC TGATGGACAG
AGGATGTTGG TTTATGAGTT TGTTCCTAAC AAAACTTTGG AATATCATCT TCATGGTTAG
ACCACTTAAA AACTTTGAGT ACTAAGTTTA TTTTCTCTAA TCTATATATT CAAGAAAGTT
GTAACCTTAA TTTGTTGTTG TAGGGAAAAA TCTTCCGGTA ATGGAGTTCT CCACTAGGTT
GCGTATCGCC TTAGGTGCTG CGAAAGGACT CGCTTACCTT CACGAAGACT GTAAGTTTTA
ACATTCACCA TTCTCATTTT CTTAACCAAG TTGCATAAAA CAGAGAAAGC TCTGTCTCTG
ACTAGTGTTA TCTTTTTGGC TGAGAAAATG GTGCAGGCCA TCCTCGGATC ATTCACCGCG
ACATCAAGTC TGCAAATATT CTCTTGGACT TCAACTTTGA TGCTATGGTG ATAAACTAGT
AGCTTGCATT CATCTACGGT TTTTTGTTAA GACTACATTG ATGACATTTT GCATTTGTTT
ATTCAGGTGG CTGATTTTGG ATTAGCTAAG TTAACATCTG ATAACAACAC TCATGTATCT
ACTCGTGTGA TGGGAACTTT CGGGTAAGCG TTTTACCGTA TGATAAGATT GTTCGTGACA
CTCAAGAAAC ATAACCTTTG TAGACTAATC TACTTTGCTT CTTTCCACAA ACATGTGTAG
ATATCTAGCT CCAGAATATG CTTCAAGCGG TAAATTAACC GAGAAATCCG ATGTTTTCTC
TTACGGAGTT ATGTTATTGG AACTTATAAC TGGAAAACGA CCGGTTGATA ATAGCATCAC
CATGGACGAC ACCTTAGTAG ATTGGGTATT CATGCATGTA ACATATGTAT CGTGTATATA
TGTTTTTCGC CTTTTTCGCG TACTAATGAT CATGAATACA GGCTCGGCCT CTTATGGCTC
GCGCGCTAGA AGATGGAAAC TTTAATGAGC TCGCAGATGC GAGGCTTGAA GGCAACTACA
ACCCGCAAGA AATGGCTCGA ATGGTGACTT GTGCCGCTGC TAGCATTCGT CATTCGGGGC
GTAAACGTCC AAAGATGAGC CAGGTGAATC AAAATTATAA CTAAAAGTCT ATTTTTGTCA
GAGAATAACA AACAAATGTT GTGGTTTTCA GATAGTAAGA GCGTTAGAAG GAGAAGTGTC
CTTAGATGCT TTAAACGAAG GTGTGAAGCC AGGACACAGT AACGTTTACG GGTCATTGGG
AGCAAGCTCG GATTATAGTC AGACATCTTA CAATGCAGAC ATGAAGAAAT TCAGACAGAT
AGCTTTGTCG AGCCAAGAAT TCCCAGTCAG TGACTGTGAA GGAACATCTA GTAATGATTC
TAGAGATATG GGAACTAAGA GCCCTACTCC TCCAAAATGA GATCGAATCA ATGATTCTGT
```

Figure 13 B.

```
   1  M   A   S   S   P   E   S   A   P   P   T   N   S   T   S   S   P   S   P   P     20
   1  ATG GCT TCT TCT CCT GAA TCT GCT CCT CCA ACA AAC TCC ACC TCT TCT CCA TCT CCA CCG    60
  21  S   N   T   N   S   T   T   S   S   P   P   A   P   S   P   P   S   P   T   P     40
  61  TCT AAT ACC AAT TCA ACC ACC TCT TCT CCG CCG GCT CCG TCT CCT CCT TCT CCT ACA CCT   120
  41  P   Q   G   D   S   S   S   S   P   P   P   D   S   T   S   P   P   A   P   Q     60
 121  CCT CAA GGA GAC TCA TCA TCA TCG CCA CCT CCT GAT TCC ACA TCT CCA CCA GCT CCA CAA   180
  61  A   P   N   P   P   N   S   S   N   N   S   P   S   P   P   S   Q   G   G   G     80
 181  GCT CCT AAC CCT CCT AAT TCC TCT AAT AAC TCT CCT TCC CCT CCG TCA CAG GGC GGT GGA   240
  81  G   E   R   G   N   G   G   N   N   G   G   N   D   T   P   P   S   R   G   S    100
 241  GGA GAA AGA GGA AAT GGA GGA AAC AAT GGT GGC AAT GAT ACT CCA CCG TCA CGC GGC TCT   300
 101  P   P   S   P   P   S   R   S   N   G   D   N   G   G   S   R   S   S   P   P    120
 301  CCT CCT TCT CCT CCT TCT AGG AGT AAT GGA GAT AAT GGT GGT AGC AGA TCA TCG CCA CCA   360
 121  G   D   T   G   G   S   R   S   D   N   P   P   S   S   G   G   S   S   G   G    140
 361  GGA GAC ACT GGA GGC TCT CGC TCA GAC AAC CCT CCT TCT AGC GGA GGA AGC AGT GGA GGA   420
 141  G   G   G   G   R   S   N   T   N   T   A   I   I   V   G   V   L   V   G   A    160
 421  GGT GGA GGT GGA AGA AGT AAT ACG AAT ACA GCG ATC ATA GTT GGT GTA TTA GTC GGA GCT   480
 161  G   L   L   M   I   V   L   I   I   V   C   L   R   R   K   K   K   R   K   D    180
 481  GGA CTT TTG ATG ATC GTT CTT ATT ATT GTG TGT CTT AGA CGC AAA AAG AAG AGA AAA GAC   540
 181  S   F   Y   P   E   P   M   K   G   N   Q   Y   Q   Y   Y   G   N   N   N   N    200
 541  TCC TTC TAC CCT GAA CCC ATG AAA GGA AAC CAA TAT CAA TAC TAT GGA AAC AAC AAC AAC   600
 201  N   N   A   S   Q   N   Y   P   N   W   H   L   N   S   Q   G   Q   N   Q   Q    220
 601  AAC AAT GCT TCA CAG AAT TAT CCG AAT TGG CAC CTA AAT TCA CAA GGC CAA AAC CAA CAA   660
 221  S   T   G   G   W   G   G   G   G   P   S   P   P   P   P   P   R   M   P   T    240
 661  TCT ACT GGT GGT TGG GGA GGC GGT GGA CCA TCA CCG CCT CCT CCT CCG CGG ATG CCT ACA   720
 241  S   G   E   D   S   S   M   Y   S   G   P   S   R   P   V   L   P   P   P   S    260
 721  AGC GGA GAA GAT TCT TCC ATG TAC TCA GGC CCA TCA CGC CCA GTT TTA CCT CCT CCT TCG   780
 261  P   A   L   A   L   G   F   N   K   S   T   F   T   Y   Q   E   L   A   A   A    280
 781  CCT GCT CTA GCC CTC GGA TTC AAC AAG AGC ACT TTT ACT TAC CAA GAG CTT GCG GCT GCA   840
 281  T   G   G   F   T   D   A   N   L   L   G   Q   G   G   F   G   Y   V   H   K    300
 841  ACA GGA GGG TTT ACG GAT GCT AAC CTT TTG GGA CAG GGA GGA TTT GGG TAT GTC CAT AAA   900
 301  G   V   L   P   S   G   K   E   V   A   V   K   S   L   K   A   G   S   G   Q    320
 901  GGA GTC TTG CCT AGC GGG AAA GAA GTA GCA GTT AAG AGT TTA AAA GCG GGT AGC GGA CAA   960
 321  G   E   R   E   F   Q   A   E   V   D   I   I   S   R   V   H   H   R   Y   L    340
 961  GGA GAG AGG GAG TTT CAA GCT GAG GTC GAT ATC ATT AGC CGT GTG CAT CAT CGG TAT CTT  1020
 341  V   S   L   V   G   Y   C   I   A   D   G   Q   R   M   L   V   Y   E   F   V    360
1021  GTT TCT TTG GTT GGA TAT TGC ATA GCT GAT GGA CAG AGG ATG TTG GTT TAT GAG TTT GTT  1080
 361  P   N   K   T   L   E   Y   H   L   H   G   K   N   L   P   V   M   E   F   S    380
1081  CCT AAC AAA ACT TTG GAA TAT CAT CTT CAT GGG AAA AAT CTT CCG GTA ATG GAG TTC TCC  1140
 381  T   R   L   R   I   A   L   G   A   A   K   G   L   A   Y   L   H   E   D   C    400
1141  ACT AGG TTG CGT ATC GCC TTA GGT GCT GCG AAA GGA CTC GCT TAC CTT CAC GAA GAC TGC  1200
 401  H   P   R   I   I   H   R   D   I   K   S   A   N   I   L   L   D   F   N   F    420
1201  CAT CCT CGG ATC ATT CAC CGC GAC ATC AAG TCT GCA AAT ATT CTC TTG GAC TTC AAC TTT  1260
 421  D   A   M   V   A   D   F   G   L   A   K   L   T   S   D   N   N   T   H   V    440
1261  GAT GCT ATG GTG GCT GAT TTT GGA TTA GCT AAG TTA ACA TCT GAT AAC AAC ACT CAT GTA  1320
 441  S   T   R   V   M   G   T   F   G   Y   L   A   P   E   Y   A   S   S   G   K    460
1321  TCT ACT CGT GTG ATG GGA ACT TTC GGA TAT CTA GCT CCA GAA TAT GCT TCA AGC GGT AAA  1380
 461  L   T   E   K   S   D   V   F   S   Y   G   V   M   L   L   E   L   I   T   G    480
1381  TTA ACC GAG AAA TCC GAT GTT TTC TCT TAC GGA GTT ATG TTA TTG GAA CTT ATA ACT GGA  1440
 481  K   R   P   V   D   N   S   I   T   M   D   D   T   L   V   D   W   A   R   P    500
1441  AAA CGA CCG GTT GAT AAT AGC ATC ACC ATG GAC GAC ACC TTA GTA GAT TGG GCT CGG CCT  1500
 501  L   M   A   R   A   L   E   D   G   N   F   N   E   L   A   D   A   R   L   E    520
1501  CTT ATG GCT CGC GCG CTA GAA GAT GGA AAC TTT AAT GAG CTC GCA GAT GCG AGG CTT GAA  1560
 521  G   N   Y   N   P   Q   E   M   A   R   M   V   T   C   A   A   A   S   I   R    540
1561  GGC AAC TAC AAC CCG CAA GAA ATG GCT CGA ATG GTG ACT TGT GCC GCT GCT AGC ATT CGT  1620
 541  H   S   G   R   K   R   P   K   M   S   Q   I   V   R   A   L   E   G   E   V    560
1621  CAT TCG GGG CGT AAA CGT CCA AAG ATG AGC CAG ATA GTA AGA GCG TTA GAA GGA GAA GTG  1680
 561  S   L   D   A   L   N   E   G   V   K   P   G   H   S   N   V   Y   G   S   L    580
1681  TCC TTA GAT GCT TTA AAC GAA GGT GTG AAG CCA GGA CAC AGT AAC GTT TAC GGG TCA TTG  1740
 581  G   A   S   S   D   Y   S   Q   T   S   Y   N   A   D   M   K   K   F   R   Q    600
1741  GGA GCA AGC TCG GAT TAT AGT CAG ACA TCT TAC AAT GCA GAC ATG AAG AAA TTC AGA CAG  1800
 601  I   A   L   S   S   Q   E   F   P   V   S   D   C   E   G   T   S   S   N   D    620
1801  ATA GCT TTG TCG AGC CAA GAA TTC CCA GTC AGT GAC TGT GAA GGA ACA TCT AGT AAT GAT  1860
 621  S   R   D   M   G   T   K   S   P   T   P   P   K   *                            634
1861  TCT AGA GAT ATG GGA ACT AAG AGC CCT ACT CCT CCA AAA TGA                          1902
```

Figure 14 A.

TCCACCGTTT GAGAAACCCT AATAACAACA TTCAAAATGG CGGACTCACC GGTGGATTCA
TCTCCTGCCC CTGAAACCTC AAATGGGACA CCACCGTCAA ACGGAACATC GCCGTCTAAT
GAGTCATCGC CGCCAACACC ACCTTCTTCA CCACCACCAT CATCAATATC TGCTCCTCCG
CCAGATATCT CCGCTTCTTT TTCACCGCCG CCTGCACCAC CAACGCAAGA AACGTCACCT
CCTACATCTC CGTCCTCATC GCCGCCTGTT GTAGCTAATC CGTCACCGCA GACTCCAGAG
AATCCTTCTC CACCTGCACC TGAAGGCTCA ACTCCTGTAA CGCCACCTGC ACCACCACAA
ACACCGTCGA ACCAATCACC GGAAAGACCA ACTCCTCCTT CTCCTGGTGC CAATGATGAC
CGAAACAGAA CCAATGGCGG AAACAACAAC AGAGACGGCT CCACACCATC ACCACCGTCG
TCAGGGAACA GAACTTCCGG TGACGGTGGC TCACCTTCAC CACCTCGGTC GATAAGCCCT
CCTCAGAATA GTGGAGATTC AGACTCATCA TCGGGTAATC ATCCACAAGC CAACATTGGA
TTGATTATTG GAGTCCTTGT AGGAGCAGGG CTTTTGCTTC TACTTGCAGT GTGTATTTGC
ATCTGTTGCA ACAGGAAGAA GAAGAAGAAA TCTCCTCAGG TCAACCACAT GCACTACTAC
AATAACAATC CTTATGGAGG AGCACCCTCA GGTAATTACA GTTTAGTATA ACTGGAATTT
AATTTGTAGC CTAATGGTGT TTGATTAGGT TTCAGAACGA TCATAGTCTA ATGGTTTCTG
CTAGCTCCAT ATGGCAAAAG GATTAGATTT ATAAGCTAAA GGAGATGTTG CATAGTGTAG
GTAATGGTGG TTATTACAAG GGAACACCTC AAGATCATGT GGTGAATATG GCTGGTCAAG
GAGGTGGGAA TTGGGGTCCA CAGCAACCTG TGTCTGGTCC TCACAGTGAT GCTTCCAACT
TAACCGGTCG AACTGCTATA CCGTCACCTC AAGCTGCAAC TCTTGGTCAC AACCAAAGCA
CTTTCACATA CGATGAACTG TCCATTGCAA CAGAAGGTTT CGCTCAGTCA AATTTGCTAG
GACAAGGAGG ATTTGGGTAT GTTCATAAAG GAGTTCTGCC TAGTGGCAAA GAAGTTGCAG
TGAAGAGTCT TAAACTTGGA AGTGGACAAG GGGAACGCGA GTTTCAAGCA GAGGTTGATA
TCATTAGCCG TGTCCATCAT CGTCATCTCG TTTCTCTTGT TGGATATTGC ATCTCTGGTG
GTCAAAGACT TTTGGTTTAT GAGTTTATAC CTAACAACAC TCTTGAATTT CATCTTCATG
GTACATTCAT CTAACAGAAT GTTTTCTTGT ATTAACAAAA CCTTTAAGTA TGGTTTCTCT
TTAATCAGGA ACATGATTGA AATTTCAGGA AAGGGTCGTC CGGTTTTGGA TTGGCCTACA
AGAGTGAAGA TTGCATTGGG ATCAGCTAGA GGCCTTGCAT ATTTGCATGA AGACTGTAAG
AAAATCTTTA TCTCACATAT TTGCATCAGT TTCTATCTCG CTCTCTACAA TATTTGAAAG
ATTGTATATG TTACATCAAT TATAGGTCAC CCTCGCATTA TCCACAGAGA TATCAAAGCT
GCAAACATTC TTCTTGATTT CAGTTTTGAG ACCAAGGTAT GTGTGTATAT ATCGACTCTT
GTACTACTTT TACTTTCATT GTCTCTCATT TTTGTTTCCA ATCTGTGTCG ATGTGTGTAT
CAGTCTTATT GTGTAAATAT ATGCAGGTGG CAGATTTTGG ATTGGCTAAG CTATCTCAAG
ACAACTATAC TCATGTCTCC ACTCGCGTCA TGGGAACTTT TGGGTAAGCA GCTTTGTAAA
ATGTCTCAAC TCATCCACAC TTATTTAGTT TCTTTCACTT GTTTTTAACA TTTTCTTGGA
TTCAGATACT TAGCTCCAGA GTATGCATCA AGCGGAAAGT TATCCGACAA ATCTGATGTT
TTCTCATTTG GAGTAATGCT TCTTGAGCTC ATAACCGGAA GACCTCCTCT GGATCTAACT
GGAGAAATGG AAGATAGCTT GGTAGATTGG GTAAGTCGGT CCCCGCCTCT TCGGTTTACT
TGTTTAATCC CAAAACACTT TCCAAAGCAA AAACAGAAAC AAATCTTACT ATTGTTGTTG
CAGGCAAGGC CTTTGTGTTT GAAAGCAGCT CAAGATGGAG ATTACAACCA ATTGGCTGAT
CCACGTCTAG AGCTAAACTA CAGTCATCAA GAGATGGTTC AAATGGCTTC TTGTGCAGCT
GCAGCAATCA GACATTCAGC AAGAAGACGG CCTAAGATGA GCCAGGTTCA AAAACTCATA
CCACTTGTTG GTTCTATTGT TATATTTTTA CTCACAATTA ATCTTGATGA TAAATGTGAC
ATACTAATGA ATCTTGAAAC ATGTGTATGG TAAATGAAAA GATTGTACGA GCACTAGAAG
GAGATATGTC AATGGATGAT CTAAGTGAGG GAACAAGACC AGGACAAAGC ACGTACTTGA
GCCCCGGGAG CGTGAGCTCA GAGTATGACG CAAGCTCGTA CACGGCAGAC ATGAAAAAGT
TCAAGAAACT GGCGTTAGAG AATAAAGAAT ATCAAAGCAG TGAATATGGT GGAACAAGTG
AGTATGGCTT AAACCCTTCT GCTTCAAGTA GTGAAGAAAT GAATAGAGGC TCAATGAAAC
GCAATCCTCA GCTTTGAAAG AAGAGACAAC ACTTGTCATA ATATTTCAGT TTTCTTCTCT

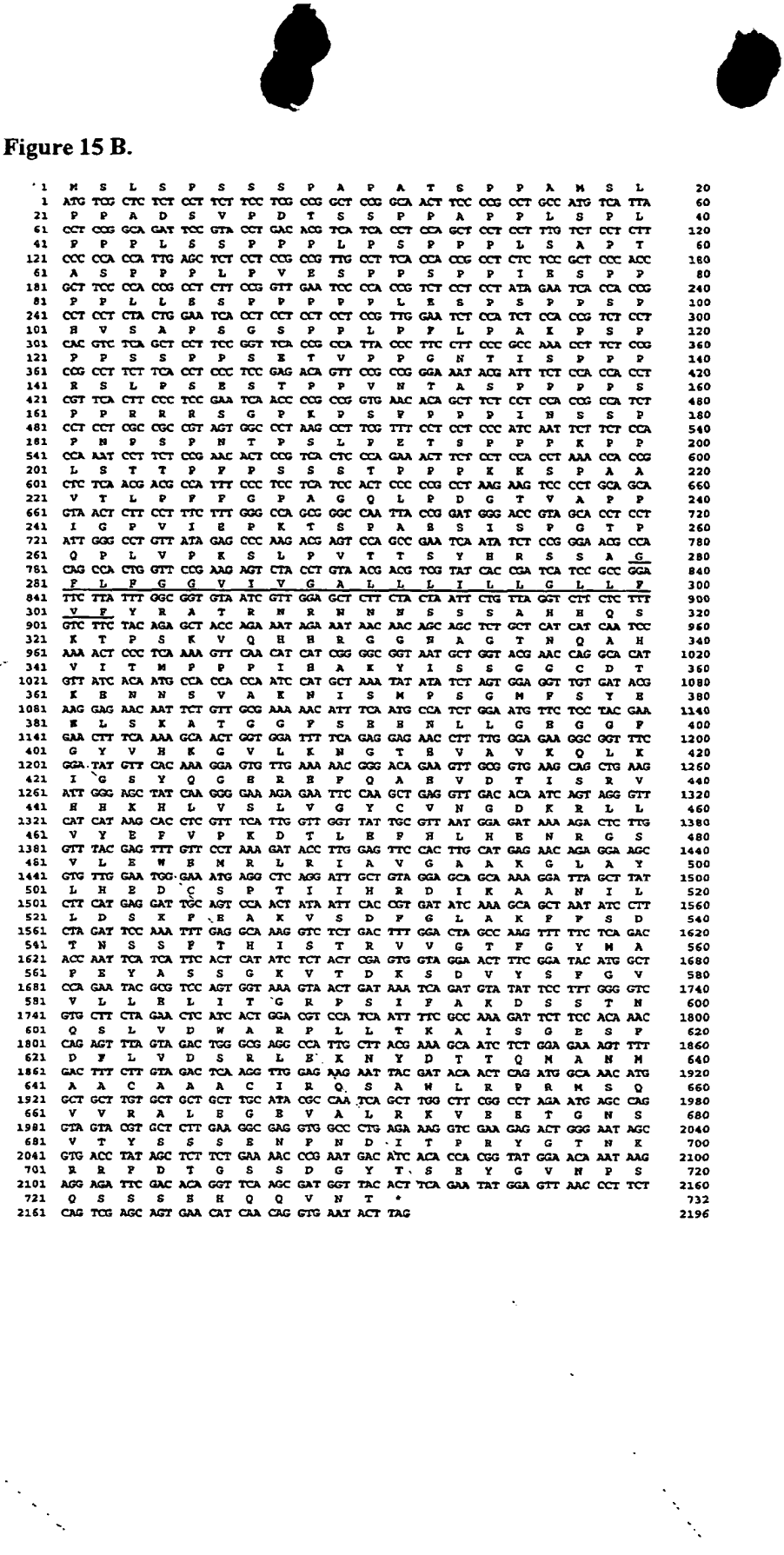
Figure 14 B.

```
   1  M  A  D  S  P  V  D  S  S  P  A  P  E  T  S  N  G  T  P  P      20
   1 ATG GCG GAC TCA CCG GTG GAT TCA TCT CCT GCC CCT GAA ACC TCA AAT GGG ACA CCA CCG      60
  21  S  N  G  T  S  P  S  N  S  S  S  P  P  T  P  P  S  S  P  P      40
  61 TCA AAC GGA ACA TCG CCG TCT AAT GAG TCA TCG CCG CCA ACA CCA CCT TCT TCA CCA CCA     120
  41  P  S  S  I  S  A  P  P  P  D  I  S  A  S  P  S  P  P  P  A      60
 121 CCA TCA TCA ATA TCT GCT CCT CCG CCA GAT ATC TCC GCT TCT TTT TCA CCG CCG CCT GCA     180
  61  P  P  T  Q  E  T  S  P  P  T  S  P  S  S  S  P  P  V  V  A      80
 181 CCA CCA ACG CAA GAA ACG TCA CCT CCT ACA TCT CCG TCC TCA TCG CCG CCT GTT GTA GCT     240
  81  N  P  S  P  Q  T  P  E  N  P  S  P  P  A  P  E  G  S  T  P     100
 241 AAT CCG TCA CCG CAG ACT CCA GAG AAT CCT TCT CCA CCT GCA CCT GAA GGC TCA ACT CCT     300
 101  V  T  P  P  A  P  P  Q  T  P  S  N  Q  S  P  E  R  P  T  P     120
 301 GTA ACG CCA CCT GCA CCA CCA CAA ACA CCG TCG AAC CAA TCA CCG GAA AGA CCA ACT CCT     360
 121  P  S  P  G  A  N  D  D  R  N  R  T  N  G  G  N  N  N  R  D     140
 361 CCT TCT CCT GGT GCC AAT GAT GAC CGA AAC AGA ACC AAT GGC GGA AAC AAC AAC AGA GAC     420
 141  G  S  T  P  S  P  P  S  S  G  N  R  T  S  G  D  G  G  S  P     160
 421 GGC TCC ACA CCA TCA CCA CCG TCG TCA GGG AAC AGA ACT TCC GGT GAC GGT GGC TCA CCT     480
 161  S  P  P  R  S  I  S  P  P  Q  N  S  G  D  S  D  S  S  S  G     180
 481 TCA CCA CCT CGG TCG ATA AGC CCT CCT CAG AAT AGT GGA GAT TCA GAC TCA TCA TCG GGG     540
 181  L  L  L  L  L  A  V  C  I  C  I  C  C  N  R  K  K  K  K  K     200
 541 CTT TTG CTT CTA CTT GCA GTG TGT ATT TGC ATC TGT TGC AAC AGG AAG AAG AAG AAG AAA     600
 201  S  P  Q  V  N  H  M  H  Y  Y  N  N  N  P  Y  G  G  A  P  S     220
 601 TCT CCT CAG GTC AAC CAC ATG CAC TAC TAC AAT AAC AAT CCT TAT GGA GGA GCA CCC TCA     660
 221  G  M  G  G  Y  Y  K  G  T  P  Q  D  H  V  V  N  M  A  G  Q     240
 661 GGT AAT GGT GGT TAT TAC AAG GGA ACA CCT CAA GAT CAT GTG GTG AAT ATG GCT GGT CAA     720
 241  G  G  G  N  H  G  P  Q  Q  P  V  S  G  P  H  S  D  A  S  N     260
 721 GGA GGT GGG AAT TGG GGT CCA CAG CAA CCT GTG TCT GGT CCT CAC AGT GAT GCT TCC AAC     780
 261  L  T  G  R  T  A  I  P  S  P  Q  A  A  T  L  G  H  N  Q  S     280
 781 TTA ACC GGT CGA ACT GCT ATA CCG TCA CCT CAA GCT GCA ACT CTT GGT CAC AAC CAA AGC     840
 281  T  F  T  Y  D  E  L  S  I  A  T  E  G  F  A  Q  S  N  L  L     300
 841 ACT TTC ACA TAC GAT GAA CTG TCC ATT GCA ACA GAA GGT TTC GCT CAG TCA AAT TTG CTA     900
 301  G  Q  G  G  F  G  Y  V  H  K  G  V  L  P  S  G  K  E  V  A     320
 901 GGA CAA GGA GGA TTT GGG TAT GTT CAT AAA GGA GTT CTG CCT AGT GGC AAA GAA GTT GCA     960
 321  V  K  S  L  K  L  G  S  G  Q  G  E  R  E  F  Q  A  E  V  D     340
 961 GTG AAG AGT CTT AAA CTT GGA AGT GGA CAA GGG GAA CGC GAG TTT CAA GCA GAG GTT GAT    1020
 341  I  I  S  R  V  H  H  R  H  L  V  S  L  V  G  Y  C  I  S  G     360
1021 ATC ATT AGC CGT GTC CAT CAT CGT CAT CTC GTT TCT CTT GTT GGA TAT TGC ATC TCT GGT    1080
 361  G  Q  R  L  L  V  Y  E  F  I  P  N  N  T  L  E  F  H  L  H     380
1081 GGT CAA AGA CTT TTG GTT TAT GAG TTT ATA CCT AAC AAC ACT CTT GAA TTT CAT CTT CAT    1140
 381  G  K  G  R  P  V  L  D  W  P  T  R  V  K  I  A  L  G  S  A     400
1141 GGA AAG GGT CGT CCG GTT TTG GAT TGG CCT ACA AGA GTG AAG ATT GCA TTG GGA TCA GCT    1200
 401  R  G  L  A  Y  L  H  E  D  C  K  K  I  P  I  S  H  I  C  I     420
1201 AGA GGC CTT GCA TAT TTG CAT GAA GAC TGT AAG AAA ATC TTT ATC TCA CAT ATT TGC ATC    1260
 421  S  H  P  R  I  I  H  R  D  I  K  A  A  N  I  L  L  D  F  S     440
1261 AGT CAC CCT CGC ATT ATC CAC AGA GAT ATC AAA GCT GCA AAC ATT CTT CTT GAT TTC AGT    1320
 441  F  E  T  K  V  A  D  F  G  L  A  K  L  S  Q  D  N  Y  T  H     460
1321 TTT GAG ACC AAG GTG GCA GAT TTT GGA TTG GCT AAG CTA TCT CAA GAC AAC TAT ACT CAT    1380
 461  V  S  T  R  V  M  G  T  F  G  Y  L  A  P  E  Y  A  S  S  G     480
1381 GTC TCC ACT CGC GTC ATG GGA ACT TTT GGA TAC TTA GCT CCA GAG TAT GCA TCA AGC GGA    1440
 481  K  L  S  D  K  S  D  V  P  S  F  G  V  M  L  L  E  L  I  T     500
1441 AAG TTA TCC GAC AAA TCT GAT GTT TTC TCA TTT GGA GTA ATG CTT CTT GAG CTC ATA ACC    1500
 501  G  R  P  P  L  D  L  T  G  E  M  E  D  S  L  V  D  W  A  R     520
1501 GGA AGA CCT CCT CTG GAT CTA ACT GGA GAA ATG GAA GAT AGC TTG GTA GAT TGG GCA AGG    1560
 521  P  L  C  L  K  A  A  Q  D  G  D  Y  N  Q  L  A  D  P  R  L     540
1561 CCT TTG TGT TTG AAA GCA GCT CAA GAT GGA GAT TAC AAC CAA TTG GCT GAT CCA CGT CTA    1620
 541  E  L  N  Y  S  H  Q  E  M  V  Q  M  A  S  C  A  A  A  A  I     560
1621 GAG CTA AAC TAC AGT CAT CAA GAG ATG GTT CAA ATG GCT TCT TGT GCA GCT GCA GCA ATC    1680
 561  R  H  S  A  R  R  R  P  K  M  S  Q  V  Q  K  L  I  P  L  V     580
1681 AGA CAT TCA GCA AGA AGA CGG CCT AAG ATG AGC CAG GTT CAA AAA CTC ATA CCA CTT GTT    1740
 581  G  S  I  I  V  R  A  L  E  G  D  M  S  M  D  D  L  S  E  G     600
1741 GGT TCT ATT ATT GTA CGA GCA CTA GAA GGA GAT ATG TCA ATG GAT GAT CTA AGT GAG GGA    1800
 601  T  R  P  G  Q  S  T  Y  L  S  P  G  S  V  S  S  E  Y  D  A     620
1801 ACA AGA CCA GGA CAA AGC ACG TAC TTG AGC CCC GGG AGC GTG AGC TCA GAG TAT GAC GCA    1860
 621  S  S  Y  T  A  D  M  K  K  F  K  K  L  A  L  E  N  K  E  Y     640
1861 AGC TCG TAC ACG GCA GAC ATG AAA AAG TTC AAG AAA CTG GCG TTA GAG AAT AAA GAA TAT    1920
 641  Q  S  S  E  Y  G  G  T  S  E  Y  G  L  N  P  S  A  S  S  S     660
1921 CAA AGC AGT GAA TAT GGT GGA ACA AGT GAG TAT GGC TTA AAC CCT TCT GCT TCA AGT AGT    1980
 661  E  E  M  N  R  G  S  M  K  R  N  P  Q  L  *                    675
1981 GAA GAA ATG AAT AGA GGC TCA ATG AAA CGC AAT CCT CAG CTT TGA                       2025
```

Figure 15 A.

CATACATCAC AAAACGGCAT CGTTTTGATG TCGCTCTCTC CTTCTTCCTC GCCGGCTCCG
GCAACTTCCC CGCCTGCCAT GTCATTACCT CCGGCAGATT CCGTACCTGA CACGTCATCA
CCTCCAGCTC CTCCTTTGTC TCCTCTTCCC CCACCATTGA GCTCTCCTCC GCCGTTGCCT
TCACCACCGC CTCTCTCCGC TCCCACCGCT TCCCCACCGC CTCTTCCGGT TGAATCCCCA
CCGTCTCCTC CTATAGAATC ACCACCGCCT CCTCTACTGG AATCACCTCC TCCTCCTCCG
TTGGAATCTC CATCTCCACC GTCTCCTCAC GTCTCAGCTC CTTCCGGTTC ACCGCCATTA
CCCTTCCTTC CCGCCAAACC TTCTCCGCCG CCTTCTTCAC CTCCCTCCGA GACAGTTCCG
CCGGGAAATA CGATTTCTCC ACCACCTCGT TCACTTCCCT CCGAATCAAC CCCGCCGGTG
AACACAGCTT CTCCTCCACC GCCATCTCCT CCTCGCCGCC GTAGTGGCCC TAAGCCTTCG
TTTCCTCCTC CCATCAATTC TTCTCCACCA AATCCTTCTC CGAACACTCC GTCACTCCCA
GAAACTTCTC CTCCACCTAA ACCACCGCTC TCAACGACGC CATTTCCCTC CTCATCCACT
CCCCCGCCTA AGAAGTCCCC TGCAGCAGTA ACTCTTCCTT TCTTTGGGCC AGCGGGCCAA
TTACCGGATG GGACCGTAGC ACCTCCTATT GGGCCTGTTA TAGAGCCCAA GACGAGTCCA
GCCGAATCAA TATCTCCGGG AACGCCACAG CCACTGGTTC CGAAGAGTCT ACCTGTAACG
ACGTCGTATC ACCGATCATC CGCCGGATTC TTATTTGGCG GTGTAATCGT TGGAGCTCTT
CTACTAATTC TGTTAGGTCT TCTCTTTGTC TTCTACAGAG CTACCAGAAA TAGAAATAAC
AACAGCAGCT CTGCTCATCA TCAATCCAAA ACTCCCTCAA AAGGTATAAA CTTTGAGATC
AATTTGTTTC AGACTGTCAC TTAAAATGTA ATGTAATTGT AAGTTATGAT CTTTTTTGCT
TGGATCAGTT CAACATCATC GGGGCGGTAA TGCTGGTACG AACCAGGCAC ATGTTATCAC
AATGCCACCA CCAATCCATG CTAAATATAT ATCTAGTGGA GGTTGTGATA CGAAGGAGAA
CAATTCTGTT GCGAAAAACA TTTCAATGCC ATCTGGAATG TTCTCCTACG AAGAACTTTC
AAAAGCAACT GGTGGATTTT CAGAGGAGAA CCTTTTGGGA GAAGGCGGTT TCGGATATGT
TCACAAAGGA GTGTTGAAAA ACGGGACAGA AGTTGCGGTG AAGCAGCTGA AGATTGGGAG
CTATCAAGGG GAAAGAGAAT TCCAAGCTGA GGTTGACACA ATCAGTAGGG TTCATCATAA
GCACCTCGTT TCATTGGTTG GTTATTGCGT TAATGGAGAT AAAAGACTCT TGGTTTACGA
GTTTGTTCCT AAAGATACCT TGGAGTTCCA CTTGCATGGT AAAATAGATA TATGATTTCA
TCCTTTTTGA TTTTGTCTTT AGTTCATTAT GTTGAGTATT GTGAGAATAT GTGTTGTAGA
GAACAGAGGA AGCGTGTTGG AATGGGAAAT GAGGCTCAGG ATTGCTGTAG GAGCAGCAAA
AGGATTAGCT TATCTTCATG AGGATTGTGA GTTGTTTCTC TTCATAATGG GAATGACAAT
GGCCTTTTTT GGGTTTTCGA CACTGATATT GATTTCTGGT GCTTGTAATG CAGGCAGTCC
AACTATAATT CACCGTGATA TCAAAGCAGC TAATATCCTT CTAGATTCCA AATTTGAGGC
AAAGGTGATC TGCTCTTTAA TCTTATCAAA GTTTGGTTTT TAGAACAGAG TTTGCCCATT
TTCGGTTTAT AGTCACACCA TTTTGTTCTA TTTAACAAGA TTACATGTAG CTTAGAGTTC
TTGTTCTCTT CAGGTCTCTG ACTTTGGACT AGCCAAGTTT TTCTCAGACA CCAATTCATC
ATTCACTCAT ATCTCTACTC GAGTGGTAGG AACTTTCGGG TAAAACACCA TCCATCCATG
CTTTATATGT TGTGTCATTG TGTTTAAAAA TATTAATTTA CGGTTTCAAG TTTCAAGTTT
CAAGTTTCAA GTTTCTTCCT TGTTGTATCG TAAATCCAGA TCAAAAGATT TATTGATTAC
TAAATGCCTT GTGCACCGTC TATTTGGTAT GCATTTAAAC AGATACATGG CTCCAGAATA
CGCGTCCAGT GGTAAAGTAA CTGATAAATC AGATGTATAT TCCTTTGGGG TCGTGCTTCT
AGAACTCATC ACTGGACGTC CATCAATTTT CGCCAAAGAT TCTTCCACAA ACCAGAGTTT
AGTAGACTGG GTAAGTCAAA GTACATGATG ATGATGATGA TACCATTAGG TTTTCTTTTT
CCCACTAGTT ATTAGATGAA AATGAAACTC CAGTATAACT GCAAGCTTGA AACTTCATTA
GACTGTAAAT TTTGATTATT CCTCAGGCGA GGCCATTGCT TACGAAAGCA ATCTCTGGAG
AAAGTTTTGA CTTTCTTGTA GACTCAAGGT TGGAGAAGAA TTACGATACA ACTCAGATGG
CAAACATGGC TGCTTGTGCT GCTGCTTGCA TACGCCAATC AGCTTGGCTT CGGCCTAGAA
TGAGCCAGGT CTGAGATTTA GTTTAAATAC ATGTATTCGT CCATATATCC GAAAGGAACA
AGCCTAATCC ATGAATACAT TTATATCTTG AAAAAAACTT GAATAGGTAG TACGTGCTCT
TGAAGGCGAG GTGGCCCTGA GAAAGGTCGA AGAGACTGGG AATAGCGTGA CCTATAGCTC
TTCTGAAAAC CCGAATGACA TCACACCACG GTATGGAACA AATAAGAGGA GATTCGACAC
AGGTTCAAGC GATGGTTACA CTTCAGAATA TGGAGTTAAC CCTTCTCAGT CGAGCAGTGA
ACATCAACAG GTGAATACTT AGTTCACAGG TTCAATAGGG CAAGTTTCAC CACAATTATT

Figure 15 B.


1 M S L S P S S S P A P A T S P P A M S L 20
1 ATG TCG CTC TCT CCT TCT TCC TCG CCG GCT CCG GCA ACT TCC CCG CCT GCC ATG TCA TTA 60
21 P P A D S V P D T S S P P A P P L S P L 40
61 CCT CCG GCA GAT TCC GTA CCT GAC ACG TCA TCA CCT CCA GCT CCT CCT TTG TCT CCT CTT 120
41 P P P L S S P P P L P S P P P L S A P T 60
121 CCC CCA CCA TTG AGC TCT CCT CCG CCG TTG CCT TCA CCA CCG CCT CTC TCC GCT CCC ACC 180
61 A S P P P L P V E S P P S P P I E S P P 80
181 GCT TCC CCA CCG CCT CTT CCG GTT GAA TCC CCA CCG TCT CCT CCT ATA GAA TCA CCA CCG 240
81 P P L L E S P P P P P L E S P S P P S P 100
241 CCT CCT CTA CTG GAA TCA CCT CCT CCT CCT CCG TTG GAA TCT CCA TCT CCA CCG TCT CCT 300
101 H V S A P S G S P P L P F L P A K P S P 120
301 CAC GTC TCA GCT CCT TCC GGT TCA CCG CCA TTA CCC TTC CTT CCC GCC AAA CCT TCT CCG 360
121 P P S S P P P S E T V P P G N T I S P P P 140
361 CCG CCT TCT TCA CCT CCC TCC GAG ACA GTT CCG CCG GGA AAT ACG ATT TCT CCA CCA CCT 420
141 R S L P S E S T P P V N T A S P P P P S 160
421 CGT TCA CTT CCC TCC GAA TCA ACC CCG CCG GTG AAC ACA GCT TCT CCT CCA CCG CCA TCT 480
161 P P R R R S G P K P S F P P P I N S S P 180
481 CCT CCT CGC CGC CGT AGT GGC CCT AAG CCT TCG TTT CCT CCT CCC ATC AAT TCT TCT CCA 540
181 P N P S P N T P S L P E T S P P P K P P 200
541 CCA AAT CCT TCT CCG AAC ACT CCG TCA CTC CCA GAA ACT TCT CCT CCA CCT AAA CCA CCG 600
201 L S T T P F P S S S T P P P K K S P A A 220
601 CTC TCA ACG ACG CCA TTT CCC TCC TCA TCC ACT CCC CCG CCT AAG AAG TCC CCT GCA GCA 660
221 V T L P F P G P A G Q L P D G T V A P P 240
661 GTA ACT CTT CCT TTC TTT GGG CCA GCG GGC CAA TTA CCG GAT GGG ACC GTA GCA CCT CCT 720
241 I G P V I E P K T S P A E S I S P G T P 260
721 ATT GGG CCT GTT ATA GAG CCC AAG ACG AGT CCA GCC GAA TCA ATA TCT CCG GGA ACG CCA 780
261 Q P L V P K S L P V T T S Y H R S S A G 280
781 CAG CCA CTG GTT CCG AAG AGT CTA CCT GTA ACG ACG TCG TAT CAC CGA TCA TCC GCC GGA 840
281 F L F G G V I V G A L L L I L L G L L F 300
841 TTC TTA TTT GGC GGT GTA ATC GTT GGA GCT CTT CTA CTA ATT CTG TTA GGT CTT CTC TTT 900
301 V F Y R A T R N R N N N S S S A H E Q S 320
901 GTC TTC TAC AGA GCT ACC AGA AAT AGA AAT AAC AAC AGC AGC TCT GCT CAT CAT CAA TCC 960
321 K T P S K V Q H H R G G N A G T N Q A H 340
961 AAA ACT CCC TCA AAA GTT CAA CAT CAT CGG GGC GGT AAT GCT GGT ACG AAC CAG GCA CAT 1020
341 V I T M P P P I E A K Y I S S G G C D T 360
1021 GTT ATC ACA ATG CCA CCA CCA ATC CAT GCT AAA TAT ATA TCT AGT GGA GGT TGT GAT ACG 1080
361 K E N N S V A K N I S M P S G M F S Y E 380
1081 AAG GAG AAC AAT TCT GTT GCG AAA AAC ATT TCA ATG CCA TCT GGA ATG TTC TCC TAC GAA 1140
381 E L S K A T G G F S E E N L L G E G G F 400
1141 GAA CTT TCA AAA GCA ACT GGT GGA TTT TCA GAG GAG AAC CTT TTG GGA GAA GGC GGT TTC 1200
401 G Y V H K G V L K N G T E V A V K Q L K 420
1201 GGA TAT GTT CAC AAA GGA GTG TTG AAA AAC GGG ACA GAA GTT GCG GTG AAG CAG CTG AAG 1260
421 I G S Y Q G E R E F Q A E V D T I S R V 440
1261 ATT GGG AGC TAT CAA GGG GAA AGA GAA TTC CAA GCT GAG GTT GAC ACA ATC AGT AGG GTT 1320
441 H H K H L V S L V G Y C V N G D K R L L 460
1321 CAT CAT AAG CAC CTC GTT TCA TTG GTT GGT TAT TGC GTT AAT GGA GAT AAA AGA CTC TTG 1380
461 V Y E F V P K D T L E F H L H E N R G S 480
1381 GTT TAC GAG TTT GTT CCT AAA GAT ACC TTG GAG TTC CAC TTG CAT GAG AAC AGA GGA AGC 1440
481 V L E W E M R L R I A V G A A K G L A Y 500
1441 GTG TTG GAA TGG GAA ATG AGG CTC AGG ATT GCT GTA GGA GCA GCA AAA GGA TTA GCT TAT 1500
501 L H E D C S P T I I H R D I K A A N I L 520
1501 CTT CAT GAG GAT TGC AGT CCA ACT ATA ATT CAC CGT GAT ATC AAA GCA GCT AAT ATC CTT 1560
521 L D S K F E A K V S D F G L A K F F S D 540
1561 CTA GAT TCC AAA TTT GAG GCA AAG GTC TCT GAC TTT GGA CTA GCC AAG TTT TTC TCA GAC 1620
541 T N S S F T H I S T R V V G T F G Y M A 560
1621 ACC AAT TCA TCA TTC ACT CAT ATC TCT ACT CGA GTG GTA GGA ACT TTC GGA TAC ATG GCT 1680
561 P E Y A S S G K V T D K S D V Y S F G V 580
1681 CCA GAA TAC GCG TCC AGT GGT AAA GTA ACT GAT AAA TCA GAT GTA TAT TCC TTT GGG GTC 1740
581 V L L E L I T G R P S I F A K D S S T N 600
1741 GTG CTT CTA GAA CTC ATC ACT GGA CGT CCA TCA ATT TTC GCC AAA GAT TCT TCC ACA AAC 1800
601 Q S L V D W A R P L L T K A I S G E S F 620
1801 CAG AGT TTA GTA GAC TGG GCG AGG CCA TTG CTT ACG AAA GCA ATC TCT GGA GAA AGT TTT 1860
621 D F L V D S R L E K N Y D T T Q M A N M 640
1861 GAC TTT CTT GTA GAC TCA AGG TTG GAG AAG AAT TAC GAT ACA ACT CAG ATG GCA AAC ATG 1920
641 A A C A A A C I R Q S A W L R P R M S Q 660
1921 GCT GCT TGT GCT GCT GCT TGC ATA CGC CAA TCA GCT TGG CTT CGG CCT AGA ATG AGC CAG 1980
661 V V R A L E G E V A L R K V E E T G N S 680
1981 GTA GTA CGT GCT CTT GAA GGC GAG GTG GCC CTG AGA AAG GTC GAA GAG ACT GGG AAT AGC 2040
681 V T Y S S S E N P N D I T P R Y G T N K 700
2041 GTG ACC TAT AGC TCT TCT GAA AAC CCG AAT GAC ATC ACA CCA CGG TAT GGA ACA AAT AAG 2100
701 R R P D T G S S D G Y T S R Y G V N P S 720
2101 AGG AGA TTC GAC ACA GGT TCA AGC GAT GGT TAC ACT TCA GAA TAT GGA GTT AAC CCT TCT 2160
721 Q S S S E H Q Q V N T * 732
2161 CAG TCG AGC AGT GAA CAT CAA CAG GTG AAT ACT TAG 2196